2021

Causal inference for the treatment of multidrug-resistant tuberculosis

BOSTON UNIVERSITY

SCHOOL OF PUBLIC HEALTH

Dissertation

CAUSAL INFERENCE FOR THE TREATMENT OF MULTIDRUG-RESISTANT TUBERCULOSIS

by

CARLY ALICIA RODRIGUEZ

B.S., Florida State University, 2012
M.P.H., Boston University, 2014

Submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

2021

Approved by

First Reader

C. Robert Horsburgh, Jr., M.D.
Professor of Epidemiology
Professor of Global Health
Professor of Biostatistics
Boston University, School of Public Health

Professor of Medicine
Boston University, School of Medicine

Second Reader

Molly F. Franke, Sc.D.
Associate Professor of Global Health and Social Medicine
Harvard Medical School

Third Reader

Carole D. Mitnick, Sc.D.
Professor of Global Health and Social Medicine
Harvard Medical School

Fourth Reader

Sara Lodi, Ph.D.
Assistant Professor of Biostatistics

ACKNOWLEDGMENTS

Foremost, I would like to thank my committee for supporting me through my doctoral studies. These exceptional professors and researchers include Charles Robert Horsburgh, Molly Franke, Carole Mitnick, and Sara Lodi. I would also like to thank Leo Martinez for his thoughtful external review of this work.

Without support from my cohort and fellow PhD students—Alyssa Harlow, Tammy Jiang, Meghan Smith, Rebecca Song, Sydney Willis, and Karla Therese Sy—I would not have been able to get through the grueling years of coursework, qualifying exams, and slog of dissertating. These women provided me with invaluable daily encouragement and friendship. I am grateful for my late grandmother, Margaret Kenney, for fostering in me a love of education and learning from a young age and my husband for cheering me on to pursue my dreams and providing immeasurable support in achieving them.

I am immensely grateful for the patients who participated in the endTB observational study, the clinicians and program staff of participating national tuberculosis programs, and endTB consortium partners including Partners In Health, Médecins Sans Frontières, and Interactive Research and Development for leading the effort to provide thousands of patients with lifesaving tuberculosis treatment.

This work was supported by the PhRMA Foundation Pre-doctoral Fellowship in Health Outcomes Research and the National Institute Of Allergy And Infectious Diseases of the National Institutes of Health under Award Number F31AI157333. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

CAUSAL INFERENCE FOR THE TREATMENT OF

MULTIDRUG-RESISTANT TUBERCULOSIS

CARLY ALICIA RODRIGUEZ

Boston University School of Public Health, 2021

Major Professor: C. Robert Horsburgh, Jr., M.D., Professor of Epidemiology, Professor of Global Health, Professor of Biostatistics, Boston University, School of Public Health; Professor of Medicine, Boston University, School of Medicine

ABSTRACT

The majority of evidence used by the World Health Organization (WHO) to inform guidelines for the treatment of multidrug-resistant tuberculosis (MDR-TB) is based on findings from observational cohort studies. Observational cohort studies have important limitations when interpreting estimates as causal effects, in contrast to randomized controlled trials (RCTs), the gold standard for assessing efficacy. Specifically, observational cohort studies are at greater risk of common threats to validity, such as selection bias and confounding. The consequences of residual bias in observational cohort studies of MDR-TB patients would be substantial, given WHO guidelines inform the treatment approach for the 500,000 patients estimated to fall sick with MDR-TB globally each year.

The goals of this dissertation are to: assess the comparative effectiveness of adding delamanid to MDR-TB regimens (Aim 1), scrutinize the potential for selection bias when using different approaches to defining the subcohort of MDR-TB patients eligible for studies using sputum culture conversion outcomes (Aim 2), and evaluate differences in the interpretation and estimates of the comparative effectiveness of adding delamanid (as

explored in Aim 1) when the time-varying nature of MDR-TB treatment is and is not accounted for (Aim 3).

In all aims, we use data from the observational cohort of the endTB initiative. The endTB initiative was launched in 2015 to rapidly expand access to two new drugs for MDR-TB, bedaquiline and delamanid, for over 2,700 patients in 17 countries. In partnership with national TB programs, a consortium of non-governmental organizations leads the initiative: Partners In Health, Médecins Sans Frontières, and Interactive Research and Development. Participants are treated in accordance with guidelines of WHO and their respective countries under routine programmatic conditions. Study activities are directed by a common protocol, data are collected using standardized forms, and adverse events (AE) are monitored through a unified pharmacovigilance system, which facilitates data consistency across sites.

In Aim 1, we investigate whether adding delamanid to MDR-TB regimens comprised of three drugs likely to be effective improves two- and six-month culture conversion. We apply a censoring approach using inverse probability weighting that accounts for MDR-TB regimen changes over the course of treatment to estimate the observational analogue of the per-protocol effect. We did not identify a difference in two- or six-month culture conversion between participants with delamanid added to their regimen and participants without delamanid. We hypothesize that delamanid did not provide a contribution to effectiveness because regimens already contained multiple efficacious drugs (e.g. linezolid, moxifloxacin/levofloxacin, bedaquiline).

In Aim 2, we used simulated data and data from the endTB observational cohort to evaluate whether extending the allowable baseline sputum culture collection interval past treatment initiation is a source of selection bias in studies using culture conversion outcomes. Two of the most influential factors that increased bias were the proportion of the cohort with a missing pre-treatment culture and the occurrence of death and loss to follow up (LTFU) in this group. These occurred infrequently in the endTB observational cohort; thus we did not observe meaningful differences when the baseline culture definition was extended past treatment initiation. In cohorts with an excess of missing pre-treatment culture data and early non-conversion events such as death and LTFU, extending the allowable interval past treatment initiation may introduce bias. Investigators should scrutinize whether extending the baseline sputum culture collection interval will inadvertently exclude these patients who may have been eligible for inclusion had they had pre-treatment sputum culture data.

In Aim 3, we used the clinical research question from Aim 1 to investigate whether implementing inverse probability of censoring weights to account for the time-varying nature of MDR-TB treatment generated results that were different from those generated through baseline-adjusted analytic approaches. Results were similar using the two types of approaches. This similarity is likely a consequence of the relative modest frequency of regimen changes and the distribution of participants with regimen changes across exposure groups and the outcome. We hypothesize that estimates may differ meaningfully for research questions where treatment changes are highly concentrated in one exposure group and these treatment changes are highly associated with the outcome.

TABLE OF CONTENTS

ACKNOWLEDGMENTS iv

ABSTRACT v

TABLE OF CONTENTS viii

LIST OF TABLES xii

LIST OF FIGURES xiv

LIST OF ABBREVIATIONS xvi

1 INTRODUCTION 1

2 COMPARATIVE EFFECTIVENESS OF ADDING DELAMANID TO MULTIDRUG-RESISTANT TUBERCULOSIS REGIMENS CONTAINING THREE DRUGS LIKELY TO BE EFFECTIVE 3

2.1 Introduction 3

2.2 Methods 5

2.2.1 Data source and study population 5

2.2.2 Design of comparative effectiveness analysis 5

2.2.3 Outcome 6

2.2.4 Definitions 6

2.2.5 Statistical analysis 7

2.2.5.1 Inverse probability of censoring weights 8

2.2.5.2 Estimating the per-protocol analogue relative risk and risk difference of culture conversion 10

2.2.5.3 Sensitivity analyses 11

2.2.6 Ethical approval *12*

2.3 Results 12

2.3.1 Analysis cohort *12*

2.3.2 Participant characteristics *12*

2.3.3 Baseline treatment regimens and regimen changes *13*

2.3.4 Two-month sputum culture conversion *14*

2.3.5 Six-month sputum culture conversion *14*

2.3.6 Sensitivity analyses *15*

2.4 Discussion 15

2.5 Tables and Figures 19

3 SELECTION BIAS IN MULTIDRUG-RESISTANT TUBERCULOSIS COHORT STUDIES USING ASSESSING SPUTUM CULTURE CONVERSION 28

3.1 Introduction 28

3.2 Methods 30

3.2.1 Absolute proportion of patients with sputum-culture conversion *30*

3.2.1.1 Quantifying bias using simulated data 30

3.2.1.2 Quantifying maximum bias in the endTB observational cohort 32

3.2.2 Relative proportion of patients with sputum-culture conversion *34*

3.3 Results 34

3.3.1 Absolute proportion of patients with sputum-culture conversion *34*

3.3.1.1 Quantifying bias using simulated data 34

3.3.1.2 Quantifying maximum bias in the endTB observational cohort data 36

3.3.2 Relative proportion of patients with sputum-culture conversion 38

3.3.2.1 endTB observational cohort data 38

3.4 Discussion 38

3.5 Tables and Figures 45

4 COMPARATIVE EFFECTIVENESS STUDIES OF TIME-VARYING TREATMENTS IN MULTIDRUG-RESISTANT TUBERCULOSIS TREATMENT COHORTS 66

4.1 Introduction 66

4.2 Methods 69

4.2.1 Data source, study population, and definitions.....69

4.2.2 Marginal versus conditional effects 69

4.2.2.1 Descriptive analysis of the distribution of censoring..... 69

4.2.3 Inverse probability censoring weighted analysis of the per-protocol analogue .70

4.2.3.1 Marginal effects of the censoring weighted per-protocol analogue..... 71

4.2.3.2 Conditional effects of the censoring weighted per-protocol analogue 72

4.2.4 Baseline-adjusted analysis of the intention-to-treat analogue.....72

4.2.4.1 Marginal effects of the baseline-adjusted intention-to-treat analogue..... 73

4.2.4.2 Conditional effects of the baseline-adjusted intention-to-treat analogue ... 73

4.3 Results 74

4.3.1 Inverse probability censoring weighted analysis of the per-protocol analogue. 74

4.3.1.1 Distribution of censoring 74

4.3.1.2 Marginal effects of the censoring weighted per-protocol analogue..... 74

4.3.1.3 Conditional effects of the censoring weighted per-protocol analogue 75

4.3.1.4 Comparison of weighted and unweighted models to assess impact of selection bias..... 75

4.3.2 Baseline-adjusted intention-to-treat analogue.....75

4.3.2.1 Marginal effects of the baseline-adjusted intention-to-treat analogue..... 75

4.3.2.2 Conditional effects of the baseline-adjusted intention-to-treat analogue ... 76

4.3.3 Observational analogue of the intention-to-treat versus per protocol analyses. 76

4.4 Discussion 77

4.5 Tables and Figures 81

5 APPENDIX.....85

6 REFERENCES 94

7 CURRICULUM VITAE.....110

LIST OF TABLES

Table 2.1 Baseline characteristics of participants receiving three drugs likely to be effective +/- delamanid, endTB observational cohort (N=363)..... 19

Table 2.2 Baseline treatment regimens of participants receiving three drugs likely to be effective +/- delamanid, endTB observational cohort (N=363)..... 24

Table 2.3 Effect of adding delamanid to a regimen composed of three drugs likely to be effective, endTB observational cohort (N=363) 27

Table 3.1 Parameters and values, simulation study of bias due to early death and loss-to-follow up events occurring during a hypothetical post-treatment initiation sputum collection interval among participants missing a pre-treatment sputum culture 45

Table 3.2 Pre-treatment initiation culture status of participants in the endTB observational cohort (N=2790) 53

Table 3.3 Sputum culture conversion and early death and loss-to-follow up events among participants missing a sputum culture in the specified interval before (-) and after (+) treatment initiation, endTB observational cohort 54

Table 3.4 Bias and precision in the relative risk of sputum culture conversion, comparison extending the interval 30 days past treatment initiation to no extension of interval past treatment initiation 60

Table 3.5 Bias and precision in the relative risk of sputum culture conversion, comparison extending the interval 60 days past treatment initiation to no extension of interval past treatment initiation 62

Table 3.6 Bias and precision in the relative risk of sputum culture conversion, comparison extending the interval 90 days past treatment initiation to no extension of interval past treatment initiation 64

Table 4.1 Comparison of analyses that assess the effect of adding delamanid to a regimen composed of three drugs likely to be effective, endTB observational cohort (N=363) 81

Table 4.2 Proportion of participants censored, by exposure group and two- and six-month culture conversion in analysis to assess the effect of adding delamanid to a regimen composed of three drugs likely to be effective, endTB observational cohort (N=363) 83

Table 4.3 Crude association of censoring with exposure group and outcome in analysis to assess the effect of adding delamanid to a regimen composed of three drugs likely to be effective, endTB observational cohort (N=363) 84

LIST OF FIGURES

Figure 2.1 Flowchart of patients eligible for the comparative effectiveness of delamanid analysis in the endTB observational cohort 23

Figure 2.2 Flowchart of participants censored due to changing treatment strategies in the endTB observational cohort 26

Figure 3.1 Culture conversion and early death and loss-to-follow up events among participants missing a pre-treatment culture, simulation study of cohort where Pt=70%, Pt | m=50%, and C | Pt=70% 46

Figure 3.2 Culture conversion and early death and loss-to-follow up events among participants missing a pre-treatment culture, simulation study of cohort where Pt=70%, Pt | m=50%, and C | Pt=50%, 70%, or 90% 48

Figure 3.3 Culture conversion and early death and loss-to-follow up events among participants missing a pre-treatment culture, simulation study of cohort where Pt varies from 60% to 90%, Pt | m=50%, and C | Pt=50%, 70%, or 90% 49

Figure 3.4 Culture conversion and early death and loss-to-follow up events among participants missing a pre-treatment culture, simulation study of cohort where Pt=70%, Pt | m varies from 250-100% and C | Pt=50%, 70%, or 90% 51

Figure 3.5 Absolute proportion of sputum culture conversion in the endTB observational cohort, by site and allowable baseline sputum culture collection interval before (-) and after (+) treatment initiation 56

Figure 3.6 Crude relative risk of sputum culture conversion by baseline characteristics, comparison extending the interval 30 days past treatment initiation to no extension of interval past treatment initiation 57
Figure 3.7 Crude relative risk of sputum culture conversion by baseline characteristics, comparison extending the interval 60 days past treatment initiation to no extension of interval past treatment initiation 58
Figure 3.8 Crude relative risk of sputum culture conversion by baseline characteristics, comparison extending the interval 90 days past treatment initiation to no extension of interval past treatment initiation 59

LIST OF ABBREVIATIONS

AE	adverse event
BDQ	bedaquiline
BMI	body mass index
$C \mid P_{mb}$	Converted | Culture positive$_{\text{maximum bias}}$
$C \mid P_{O}$	Converted | Culture positive$_{\text{observed}}$
$C \mid P_{t}$	Converted | Culture positive$_{\text{truth}}$
Cfz	clofazimine
CI	confidence interval
Csn/Tzd	cycloserine/terizidone
Db	decibel
Dlm	delamanid
E	ethambutol
Eto/Pto	ethionamide/prothionamide
HbA1C	Hemoglobin A1C
HIV	human immunodeficiency virus
Hz	hertz
ImCil/M	imipenem-cilastatin-amoxicillin clavulanate or meropenem
ITT	Intention-to-treat
Lfx/Mfx	levofloxacin/moxifloxacin
LTFU	loss to follow up
Lzd	linezolid
m	Culture missing$_{\text{observed}}$

M.tb	*Mycobacterium tuberculosis*
MDR-TB	multidrug-resistant tuberculosis
PAS	p-aminosalicylic acid
P_t	Culture positive$_{truth}$
$P_t \mid m$	Culture positive$_{truth}$ | Culture missing$_{observed}$
QTcF	QT interval with Fridericia's correction
RCT	randomized controlled trial
RD	risk difference
RR	risk ratio
RR-TB	rifampin-resistant tuberculosis
S/A	streptomycin/amikacin
SAE	serious adverse event
SD	standard deviation
SW	stabilized weight
SW^C	stabilized censoring weight
TB	tuberculosis
W	unstabilized weight
W^A	unstabilized treatment weight
W^C	unstabilized censoring weight
WHO	World Health Organization
XDR-TB	extensively drug-resistant tuberculosis
Z	pyrazinamide

1 INTRODUCTION

An estimated 10 million people globally developed tuberculosis (TB) disease in 2019.(1) Of these, nearly 500,000 harbored isolates resistant to at least rifampin (rifampin resistant TB, RR-TB), or rifampin and isoniazid (multidrug-resistant TB, MDR-TB), the most effective drugs to treat TB.(1) Conventional treatment for MDR-TB is complex and difficult. Patients must take a regimen of at least four drugs in which the probability of resistance is low (termed "likely effective" drugs).(2) Many of these drugs result in debilitating side effects such as peripheral neuropathy and permanent hearing loss.(3,4) Conventional treatment for MDR-TB is also long, lasting up to 24 months.(5) Even after enduring extensive and toxic treatment, outcomes are often poor: only half of patients reach bacteriologic cure (i.e., sputum culture conversion) or complete treatment.(6) Outcomes are worse for patients whose infecting strains harbor additional resistance to fluoroquinolones and injectable agents (historically classified as extensively drug resistant TB, XDR-TB); only one-third of patients have a successful treatment outcome.(6)

The introduction of bedaquiline and delamanid—the first drugs with novel mechanisms of action against *M. tuberculosis* (*M.tb*) in nearly 50 years—have offered the potential to improve treatment for MDR-TB. Between 2012 and 2014, the United States (US) Food and Drug Administration and the European Medicines Agency approved bedaquiline and the European Medicines Agency approved delamanid for TB.(7,8) Recently, the WHO issued guidance revisions recommending bedaquiline as a first-line treatment for MDR-TB.(9) WHO guidelines also endorse the addition of delamanid to regimens that do not

contain ≥5 likely effective drugs. A number of RCTs for these drugs are currently in progress.(10–13) Many clinical questions regarding the use of these drugs, especially in combination, remain unanswered at present and will not be addressed by the RCTs.

Much of the evidence guiding treatment of MDR-TB, including the most recent WHO guideline revisions, has been based on observational data.(7,9,14–16) However, these studies have generally lacked longitudinal assessment of treatment and inadequately controlled for confounding, limiting their validity. With RCTs still the uncontested gold standard for assessing treatment efficacy, a body of research has emerged providing evidence that causal effects can be inferred in observational data.(17,18) The validity of these studies relies on scrutinizing common threats such as confounding and selection bias and implementing appropriate analytic methods to control for biases introduced when a treatment is not randomized. Applying methods, such as Robins generalized methods, or "g-methods,"(19) to rigorously collected longitudinal data on MDR-TB treatment can improve evidence that emerges from non-randomized studies of bedaquiline, delamanid, and other anti-TB drugs.

2 COMPARATIVE EFFECTIVENESS OF ADDING DELAMANID TO MULTIDRUG-RESISTANT TUBERCULOSIS REGIMENS CONTAINING THREE DRUGS LIKELY TO BE EFFECTIVE

2.1 Introduction

Delamanid (OPC-67683, Deltyba®) received conditional approval for use in treatment of TB by the European Medicines Agency in 2014.(8) It was only the second drug with a novel mechanism of action against *M.tb* to receive marketing authorization from a stringent regulatory authority in forty years. Results of the Phase IIb RCT showed delamanid to be efficacious: 45.4% of patients in the delamanid-plus-standard-treatment arm achieved culture conversion at 2 months vs 29.6% in the standard treatment arm (p=0.008).(20) However, a subsequent Phase III trial found no clinically relevant or statistically significant difference between arms: 87.6% in the delamanid-plus-standard-treatment arm achieved conversion by six months vs 86.1% in the placebo-plus-standard treatment arm.(21) Based on the Phase III trial results, the WHO concluded in 2018 interim guidance that "the demonstrated benefit of delamanid when added to an optimized background regimen was small."(22) Delamanid was subsequently categorized as a lower priority Group C drug that may be included in longer MDR-TB regimens if a regimen cannot be composed with the more effective Group A and Group B drugs.(9)

One observed difference between the Phase IIb and Phase III trials is composition of the optimized background regimen. Participants in the Phase II trial received 4-5 drugs on

average. The Phase III trial used an optimized background regimen comprising a mean of 6.5 drugs. Other research has elucidated that patients treated with conventional MDR-TB treatment experience maximal treatment response and lowest odds of mortality when they receive at least 5 drugs in the initial phase.(23) This raises the possibility that the divergence of findings between the two trials may be because participants in the Phase III trial received a larger number of effective drugs, potentially masking delamanid's contribution to treatment outcomes. The effect of delamanid when added to a weak regimen—such as in patients where a regimen is limited to only three drugs likely to be effective—was unknown.

Recent WHO guidelines suggest that delamanid may have a critical role in strengthening regimens and have called for studies of the drug in the context of regimens compromised by resistance or intolerability, two features that often result in patients receiving suboptimal regimens with few drugs.(16) In an early report of 66 patients receiving delamanid under compassionate use, 80% were culture negative at 6 months.(24) Patients had, on average, received only 3.3 drugs likely to be effective. This treatment response—which far exceeded that from historical cohorts treated without delamanid—suggested that delamanid may provide an advantage in terms of effectiveness among patients with limited treatment options. Subsequent descriptive studies of patients treated with delamanid showed similar early success, with 70-95% of patients achieving culture conversion within six months of delamanid initiation.(25–29) However, to date, no observational studies have directly compared the effectiveness of delamanid-containing regimens to delamanid-free regimens.

Here, we evaluated the effect of adding delamanid for 24 weeks to MDR-TB regimens comprising only three drugs likely to be effective. We estimated two-month and six-month culture conversion in a large prospective, research cohort of patients receiving delamanid.

2.2 **Methods**

2.2.1 Data source and study population

We included participants with a positive baseline culture and documented RR-TB or MDR-TB from the endTB observational cohort (NCT02754765), a prospective research cohort started in 2015 to expand access to bedaquiline and delamanid in 17 countries. Study protocol details have been reported previously.(30) In short, participants were treated under routine programmatic conditions in accordance with guidelines of their respective countries and of WHO during the study period.(14,31) Clinical care was further informed by the endTB clinical guide.(32) Parent study activities were directed by a common protocol.(30) Data were collected using standardized forms and adverse events were monitored through a unified pharmacovigilance system.(33)

2.2.2 Design of comparative effectiveness analysis

Because the intent of the present effort was to draw inferences on the comparative effectiveness of adding delamanid for 24 weeks to an MDR-TB regimen of three drugs likely to be effective, we designed our analysis using target trial emulation.(34–38) We designed a hypothetical, pragmatic, RCT—a "target trial"—to answer the causal question

of interest. We detailed criteria of the hypothetical, pragmatic RCT, including eligibility criteria, treatment strategy, treatment assignment, follow-up, outcome, causal contrast, and statistical analysis (Appendix 5.1). We then emulated this target trial with our observational data and conducted a statistical analysis to control for threats to internal validity such as selection bias and confounding.

2.2.3 Outcome

Culture conversion is used as an interim microbiological indicator and surrogate endpoint in both observational studies and RCTs.(39,40) We assessed two- and six-month culture conversion risks. We defined culture conversion as the first of two consecutive negative cultures collected at least 15 days apart. For two-month culture conversion, the first culture had to be collected within the first 56 days of treatment and the second in the first 210 days; for six-month culture conversion the first culture had to be collected within the first 180 days of treatment initiation and the second in the first 210 days. Participants who died or were LTFU before conversion were considered as not having converted. LTFU was defined as treatment interruption (i.e. no treatment) for ≥2 months.

2.2.4 Definitions

We included participants with a positive baseline sputum culture, defined as any culture on a sputum specimen collected up to 90 days before treatment initiation. We excluded patients treated in the Democratic People's Republic of Korea due to substantial differences in diagnosis, treatment delivery, and lack of HIV testing, compared to the rest of the cohort. Likely effectiveness was defined as: (1) a drug for which resistance testing

indicated that the participant's *M.tb* strain was not resistant to the drug, or (2) a drug for which no resistance testing had been conducted and the participant had not previously received the drug for one month or more. Baseline exposure was categorized as one of the following: (1) receiving a regimen of delamanid plus a background regimen of three drugs likely to be effective, (2) receiving a regimen of three drugs likely to be effective, none of which was delamanid, or (3) neither exposure group of interest. In order to allow for early adjustments to treatment regimens (likely unrelated to efficacy or safety), we categorized the baseline exposure according to the drugs prescribed on the 7th day after treatment initiation.

2.2.5 Statistical analysis

An analysis estimating the observational analogue of the intention-to-treat (ITT) effect of the target trial would include all participants, classify their treatment according to baseline, and adjust for baseline confounders. This analysis would estimate the effect of *initiating* delamanid plus a background regimen of three drugs versus a background regimen of three drugs, not including delamanid. However, we observed that during follow-up, exposure status changed: for some participants in the delamanid-containing group, delamanid was discontinued; for some participants in the delamanid-free group, it was started; and in both groups, some patients experienced changes in the number background drugs. Therefore, the ITT analysis could underestimate the effect of delamanid, if any exists. A more useful causal contrast is the observational analogue of the per-protocol effect, which estimates the effect of delamanid among participants whose baseline exposure status is maintained for the duration of the study period (24

weeks). Because MDR-TB treatment can vary over time, there is potential for time-dependent confounding. Time-dependent confounding can present itself when a time-varying variable is a risk factor for the outcome, predicts subsequent treatment, and is affected by past treatment.(19,41) This is likely the case in MDR-TB for factors such as microbiologic monitoring (e.g. sputum smear) and drug resistance monitoring.(42) We applied a simple technique for time-varying treatments whereby participants are artificially censored when their treatment deviates from that administered at baseline. To control for selection bias due to artificial censoring, we applied inverse probability of censoring weights. This technique effectively forces the time-varying treatment to be non-time varying, thus eliminating time-dependent confounding.(43)

2.2.5.1 Inverse probability of censoring weights

To simulate creation of the per-protocol population, we censored observations in the delamanid-containing group if delamanid had been discontinued for >2 consecutive weeks and there was no evidence that delamanid discontinuation was in response to an adverse event; if there was a documented adverse event, we did not censor. In the delamanid-free group, we censored observations if delamanid was added for >2 consecutive weeks. In both groups, observations were censored if any effective background drug was added or removed for >2 consecutive weeks and the resulting regimen contained either less than three drugs likely to be effective or more than three drugs likely to be effective.

Outright censoring of follow-up time after a change in exposure group will induce selection bias. To control for selection bias due to artificial censoring, for each individual

and for each week, we estimated time-varying inverse probability of censoring weights equal to the inverse of the probability of being uncensored, i.e. maintaining treatment consistent with the baseline exposure group. The use of inverse probability of censoring weighting creates the pseudopopulation that would have been observed if, counterfactually, baseline exposure was maintained in all participants (unstabilized censoring weights), or baseline exposure changes were at random (stabilized censoring weights). In the pseudopopulation, there is no confounding by baseline or time-varying confounders.

To estimate the weights, we fitted a pooled logistic regression model to estimate the probability that each participant remained on their baseline exposure (i.e. was not censored) conditional on time-varying predictors of changing treatment and time since baseline. These predictors included time-varying number of Group A drugs, sputum smear result, number of adverse events, hospitalization, time, and a quadratic function of time (Model 2, Appendix 5.2). Some missingness occurred for time-varying sputum smear result, resulting in the exclusion of participants without these data. We conducted a sensitivity analysis comparing estimates from models with and without time-varying sputum smear result (Appendix 5.2 and Appendix 5.3).

For unstabilized censoring weights (W^C), we divided 1 by the probability of a participant remaining on their baseline exposure using the formula below where: *t* denotes the time in weeks from treatment initiation to the time-fixed culture conversion endpoint (2 months, 6 months, death or LTFU), A is the baseline exposure (1=delamanid-containing, 0=delamanid-free), C(*t*) indicates artificial censoring due to switching exposure groups at

time *t* (1= censored, 0=not censored), $\bar{L}(t)$ represents the vector of time-varying covariates at time *t* and L represents the vector of baseline covariates used to model time-varying covariate history. At time *t*, the denominator of the weight is the product of the predicted probability of being uncensored.

$$W^C(t) = \prod_{k=0}^{t} \frac{1}{\Pr\left[C(k)=0 \mid L, \bar{L}(k),\ C(k-1)=0,\ A\right]}$$

For statistical efficiency of our risk estimate, we additionally derived stabilized censoring weights (SW^C) by fitting a second pooled logistic regression to estimate the probability of the participant remaining on their baseline exposure (i.e. not being censored), conditional on time, as represented by the following formula:

$$SW^C(t) = \prod_{k=0}^{t} \frac{\Pr\left(C(k)=0 \mid C(k-1)=0,\ A\right)}{\Pr\left[C(k)=0 \mid L, \bar{L}(k),\ C(k-1)=0,\ A\right]}$$

We fitted multiple models, identifying predictors of exposure group changes *a priori* based on content knowledge. Comparisons of these models and the mean and standard deviation of estimated weights are described in Appendix 5.2 and 5.3.

2.2.5.2 Estimating the per-protocol analogue relative risk and risk difference of culture conversion

We identified baseline confounders of treatment using content knowledge and directed acyclic graphs (Appendix 5.4). We fitted multiple models to control for confounding by baseline characteristics, which are detailed in Appendix 5.5. We selected a final baseline model comprised of age, sex, whether the participant was in the hospital at treatment initiation, the number of Group A drugs in the regimen, whether the patient was on

imipenem-cilastatin, body mass index <18.5, HIV infection, and hepatitis C. Missing data were rare for most confounders (<1%), with the exception of baseline cavitation on chest radiography (Table 1, 10.7% missing). Following the principles of Greenland,(44) we conducted a missing indicator analysis for a composite variable representing cavitation and smear grade, which has been found to be highly predictive of poor treatment outcome (Appendix 5.6).(45,46)

Using a weighted logistic regression model adjusted for baseline confounders of treatment, we estimated the predicted probabilities of culture conversion for each uncensored participant. We then used the mean predicted probability of conversion by exposure group to calculate the point estimate for the relative risk and risk difference. Confidence intervals were calculated using nonparametric bootstrapping with 500 samples. Because we controlled for baseline confounders of treatment in the final logistic regression model (as opposed to using a composite inverse probability of treatment and censoring weight), estimates reflect conditional effects.

2.2.5.3 Sensitivity analyses

To account for the possibility that adjustment for the number of Group A drugs did not adequately control for the efficacy of the background regimen across exposure groups and given bedaquiline's potent bactericidal activity,(47) we conducted a sensitivity analysis among patients who had received bedaquiline. We restricted the delamanid-containing group to those who also received bedaquiline. All participants in the delamanid-free group received bedaquiline; by design, all participants in the endTB observational cohort either received delamanid and/or bedaquiline.

2.2.6 Ethical approval

We obtained ethical approval from the central ethics review committees for each consortium partner and local ethical approvals in each country. Participants provided written informed consent.

2.3 **Results**

2.3.1 Analysis cohort

Between April 1, 2015 and September 30, 2018, 2755 patients were initiated on a first regimen with bedaquiline and/or delamanid and consented to participate in the endTB observational study (Figure 2.1). We excluded 1996 (72.5%) participants whose baseline regimen did not fall into an exposure group of interest. Patients who did not have RR/MDR-TB (n=6), had a negative or missing baseline sputum culture (n=358), or were treated in the Democratic People's Republic of North Korea (N=32) were excluded, leaving 363 participants (N=125 delamanid-containing, N=238 delamanid-free) in the analysis cohort (Figure 2.1).

2.3.2 Participant characteristics

The majority of patients were treated in Kazakhstan (30.4%), Georgia (16.8%), Peru (14.0%), and Pakistan (12.1%) (Table 2.1). Exposure groups were comparable with regard to age, sex, and indicators of disease severity such as cavitation, bilateral disease, and smear grade; however, missing data on cavitation and bilateral disease was more prevalent among participants in the delamanid-free group. The delamanid-containing

group consistently had a greater proportion of participants with comorbidities including HIV, Hepatitis B, Hepatitis C, malnutrition, and diabetes. Approximately 10% of participants in the delamanid-containing group had been treated exclusively with first-line anti-TB drugs, compared to <1% of participants in the delamanid-free group (Table 2.1).

2.3.3 Baseline treatment regimens and regimen changes

Although participants in both groups received a background regimen of three drugs likely to be effective, there was substantial heterogeneity in the drugs comprising regimens (Table 2.2). On average, participants in the delamanid-containing group had fewer Group A drugs, those classified as priority drugs in the 2020 WHO MDR-TB treatment guidelines.(5) While 69.8% (n=166/238) of the delamanid-free group participants received all three Group A drugs (bedaquiline, linezolid, levofloxacin/moxifloxacin) or two Group A drugs (bedaquiline, linezolid) and one Group B drug (clofazimine), only 39.2% (n=49/125) of the delamanid-containing group participants received the combination containing bedaquiline, linezolid, and clofazimine and none received all 3 Group A drugs (Table 2.2).

Baseline exposure group was maintained for 24 continuous weeks (Figure 2.2) in approximately 60% of participants. Of the 132 whose exposure group changed, 34 had short term (<2 week) changes; in one participant, delamanid was removed due to an adverse event. These participants were not censored. The remaining 97 participants had regimen adjustments that resulted in an exposure group change and censoring: in 8, delamanid was added; in 4 delamanid was withdrawn without a documented, related

adverse event; and 85 had a background regimen change resulting in a change to the participant's exposure group (Figure 2.2). Unstabilized censoring weights had a mean of 1.35 (SD 0.44); stabilized censoring weights were a mean of 0.98 (SD 0.32) (Appendix 5.2). Adjusted estimates were calculated based on the sample of 349 participants with complete data (Table 2.3). Twelve participants were excluded from the analysis due to missing data on time-varying sputum smear used to calculate censoring weights. The exclusion of this variable and subsequent inclusion of these participants in a sensitivity analysis resulted in only a minor change to the point estimate (Model 4 vs Model 2, Appendix 5.3).

2.3.4 Two-month sputum culture conversion

At two months, 49.6% of participants in the delamanid-containing group and 55.9% of participants in the delamanid-free group experienced culture conversion. Among participants who did not have culture conversion, 6 (4.8%) (3 deaths, 3 LTFU) participants in the delamanid-containing group and 7 (2.9%) (2 deaths, 5 LTFU) participants in the delamanid-free group died or were LTFU in the first two months (data not shown). The adjusted analysis using stabilized weights resulted in a risk ratio of 0.93 (95% CI: 0.70, 1.33) and a risk difference of -0.04 (95% CI: -0.18, 0.16) (Table 2.3) for delamanid-containing versus delamanid-free regimens.

2.3.5 Six-month sputum culture conversion

By six months, 80.8% of participants in the delamanid-containing and 89.1% of participants in the delamanid-free group experienced sputum culture conversion. Fifteen

(12.0%) (7 deaths, 8 LTFU) participants in the delamanid-containing group died or were LTFU before conversion versus 14 (5.9%) (4 deaths, 10 LTFU) in the delamanid-free group (data not shown). Using stabilized weights, the adjusted risk ratio of six-month sputum culture conversion was 0.93 (95% CI: 0.78, 1.05) and risk difference was -0.06 (95% CI: -0.89, 0.05) (Table 2.3) for delamanid-containing versus delamanid-free regimens.

2.3.6 Sensitivity analyses

When we restricted analyses to participants receiving bedaquiline in their baseline regimen, the two-month risk ratio of culture conversion was 0.97 (95% CI: 0.66, 1.42) and risk difference was -0.02 (95% CI: -0.20, 0.20) (Table 2.3). Estimates for six-month culture conversion shifted rightward towards the null, with a risk ratio of 1.02 (95% CI: 0.94, 1.11) and risk difference of 0.02 (95% CI: -0.06, 0.10) (Table 2.3).

2.4 **Discussion**

We identified a null effect of adding delamanid to a regimen containing only three other drugs likely to be effective. In this context, delamanid added little, consistent with the previously reported Phase III delamanid trial but different from those of the Phase IIb trial.(20,21)

Adding delamanid to a regimen with only three drugs did not improve culture conversion in our study.(5) Our treatment groups were defined by the *quantity* of drugs in a regimen. However, the efficacy of drugs in the regimen cannot be ignored. The advent of

bedaquiline, repurposed drugs, such as linezolid and clofazimine, and late generation quinolones, such as levofloxacin/moxifloxacin, have innovated the TB drug and regimen landscape. Regimens of lesser efficacy, like those used in the Phase IIb trial that showed a significant effect of delamanid,(20) are not represented in large numbers in the endTB cohort. The majority of participants in the endTB cohort received linezolid (82%), clofazimine (73%), bedaquiline (73%), or levofloxacin/moxifloxacin (58%) at baseline.(46) These regimens more accurately reflect the efficacy of those used in the Phase III trial,(21) which also identified no effect of adding delamanid to a regimen. Thus, it is not surprising our findings are similar. Whether delamanid can improve the effectiveness of regimens compromised by toxicity or resistance to the aforementioned new and repurposed drugs remains unanswered. The regimens needed to answer this research question would have efficacy that aligns more closely with the efficacy of the background regimen employed in the Phase IIb trial.

There was more death and LTFU in the delamanid-containing group (12.0%) than in the DLM-free group (5.9%). This may be because participants in the delamanid-containing group had more comorbid conditions highly predictive of early death and poor treatment outcome, such as HIV, and more fluoroquinolone and injectable resistance.(46,48) Our outcome definition considered death and LTFU occurring before conversion as non-conversion events, which is in accordance with how these events are classified as unsuccessful final treatment outcomes according to WHO criteria. An alternative approach would be to create a second set of censoring weights for death and LTFU, exclude participants with these events, and weight the remaining participants accordingly

to adjust for selection bias from artificially censoring participants who died or were LTFU.(49) The estimate from such an analysis would represent the effect of adding delamanid to the regimen, assuming there were no deaths or LTFU in the study.

Our analysis assessed culture conversion at two and six months. While culture conversion is a commonly used surrogate endpoint or prognostic marker for treatment outcome in RCTs, conversion does not perfectly predict clinical benefit. In fact, the extent to which sputum-based endpoints at various timepoints can predict treatment outcomes (e.g. favorable vs unfavorable, relapse free cure) is a point of debate. Reanalysis of trials in drug-sensitive TB and MDR-TB have provided contradictory evidence in which some studies have concluded sputum-based markers can predict outcomes,(40,50–52) while other studies provide evidence questioning the statistical validity of sputum-based markers.(53–55) Observational studies have also yielded contradictory findings across studies.(39,56,57) Despite the absence of conclusive evidence on the role of conversion in predicting treatment outcome, culture conversion remains a commonly used early outcome in MDR-TB treatment. Future analyses of the discriminatory power of two and six month conversion are needed within the context of the endTB cohort to fully understand how our study's findings extend to final treatment outcome.

Our study has several limitations. First, despite narrowly defining our exposure groups, there may be residual differences in the quality of the background regimen across exposure groups. We attempted to control for the efficacy of regimens across exposure groups by adjusting for the number of Group A drugs. Few efforts have been made to meaningfully capture the heterogeneity of MDR-TB regimens in comparative

effectiveness studies, likely because hundreds or even thousands of distinct regimens can be represented in any one cohort. For example, the 2012 individual patient data meta-analysis comprises over 9000 patients on 1626 different regimens.(58,59) Further methodologic work in this area is needed, as global treatment guidance relies largely on observational cohorts for the evidence base.(23,58) Second, there are a number of assumptions that must be met in observational studies in order for our estimates to be valid. The assumption that often garners the most focus is that all confounders must be adjusted for. We cannot rule out the possibility of residual confounding, though we built multiple baseline models to adjust for measured confounders of treatment and conversion. The probability there were additional unmeasured confounders is likely low, given that we collected a multitude of data on baseline factors. All prognostic factors predicting exposure group changes must also have been identified and accurately modeled to rule out unmeasured confounding and residual confounding. Correct model specification cannot be guaranteed, however we built multiple censoring weight models and the results did not meaningfully change. A major strength of our study and the reason we were able to conduct a censoring weighted analysis is that we collected an abundance of time-varying risk factors that have historically not been represented in previous cohorts. Without these data, such an analysis would not have been possible.

In sum, we did not identify an effectiveness benefit of adding delamanid to an MDR-TB regimen with a suboptimal number of drugs. However, critical questions related to delamanid's role in MDR-TB regimens remain, including whether delamanid can improve the effectiveness of regimens comprised of drugs with suboptimal efficacy.

2.5 Tables and Figures

Table 2.1 Baseline characteristics of participants receiving three drugs likely to be effective +/- delamanid, endTB observational cohort (N=363)

	Exposure group				Overall, N=363	
	DLM-containing, n=125		DLM-free, n=238			
	n	(%)	N	(%)	n	(%)
Demographic						
Site						
Armenia	11	(8.8)	10	(4.2)	21	(5.8)
Bangladesh	9	(7.2)	25	(10.5)	34	(9.4)
Belarus	15	(12.0)	2	(0.8)	17	(4.7)
Georgia	12	(9.6)	49	(20.6)	61	(16.8)
Haiti	5	(4.0)	0	0	5	(1.4)
Indonesia	0	0	1	(0.4)	1	(0.3)
Kazakhstan	54	(43.2)	56	(23.5)	110	(30.3)
Kyrgyzstan	2	(1.6)	2	(0.8)	4	(1.1)
Lesotho	4	(3.2)	5	(2.1)	9	(2.5)
Myanmar	3	(2.4)	0	0	3	(0.8)
Pakistan	9	(7.2)	35	(14.7)	44	(12.1)
Peru	0	0	51	(21.4)	51	(14.0)
South Africa	1	(0.8)	0	0	1	(0.3)
Vietnam	0	0	2	(0.8)	2	(0.6)
Age, median (SD)	39	(11.6)	35	(12.6)	36	(12.3)
Gender						
Male	80	(64.0)	166	(69.7)	246	(67.8)
Disease severity						
Baseline cavity						
Missing	6	(4.8)	33	(13.9)	39	(10.7)

Cavitation	90	(72.0)	140	(58.8)	230	(63.4)
No cavitation	29	(23.2)	65	(27.3)	94	(25.9)
Bilateral disease						
Missing	3	(2.4)	26	(10.9)	29	(8.0)
Bilateral	89	(71.2)	145	(60.9)	234	(64.5)
Non-bilateral	33	(26.4)	67	(28.2)	100	(27.5)
Smear grade (-90 days, +0 days from treatment initiation)						
Missing	1	(0.8)	4	(1.7)	5	(1.4)
Scanty 1-3	3	(2.4)	2	(0.8)	5	(1.4)
Scanty 4-9	4	(3.2)	8	(3.4)	12	(3.3)
One plus	44	(35.2)	76	(31.9)	120	(33.1)
Two plus	23	(18.4)	44	(18.5)	67	(18.5)
Three plus	15	(12.0)	37	(15.5)	52	(14.3)
Disease site						
Extrapulmonary	0	0	3	(1.3)	3	(0.8)
Pulmonary	125	(100.0)	235	(98.7)	360	(99.2)
Comorbidities						
HIV						
Missing	1	(0.8)	0	0	1	(0.3)
Negative	106	(84.8)	227	(95.4)	333	(91.7)
Positive	18	(14.4)	11	(4.6)	29	(8.0)
Hepatitis B						
Missing	1	(0.8)	0	0	1	(0.3)
Negative	116	(92.8)	231	(97.1)	347	(95.6)
Positive	8	(6.4)	7	(2.9)	15	(4.1)
Hepatitis C						
Missing	2	(1.6)	0	0	2	(0.6)
Negative	96	(76.8)	212	(89.1)	308	(84.8)
Positive	27	(21.6)	26	(10.9)	53	(14.6)

Malnutrition (body mass index <18.5)						
Missing	2	(1.6)	0	0	2	(0.6)
No malnutrition	71	(56.8)	155	(65.1)	226	(62.3)
Malnutrition	52	(41.6)	83	(34.9)	135	(37.2)
Anemia						
Missing	4	(3.2)	2	(0.8)	6	(1.7)
No anemia	70	(56.0)	120	(50.4)	190	(52.3)
Anemia	51	(40.8)	116	(48.7)	167	(46.0)
Diabetes						
Missing	1	(0.8)	3	(1.3)	4	(1.1)
No diabetes	97	(77.6)	208	(87.4)	305	(84.0)
Diabetes	27	(21.6)	27	(11.3)	54	(14.9)
Indications						
Previous TB treatment						
Missing	7	(5.6)	4	(1.7)	11	(3.0)
Only with first-line drugs	13	(10.4)	2	(0.8)	15	(4.1)
With second-line drugs	105	(84.0)	232	(97.5)	337	(92.8)
Indication for DLM and BDQ						
Construction of regimen of 4 likely effective second-line drugs not possible	115	(92.0)	238	(100.0)	353	(97.2)
Other high risk of unfavorable outcome	10	(8.0)	0	0	10	(2.8)
Existing neuropathy						
Missing	38	(30.4)	57	(23.9)	95	(26.2)
Existing neuropathy	20	(16.0)	36	(15.1)	56	(15.4)
None	67	(53.6)	145	(60.9)	212	(58.4)
Hearing loss at baseline (2 contiguous Hz each >= 30dB in either ear)						
Missing	52	(41.6)	81	(34.0)	133	(36.6)
At least some hearing loss	44	(35.2)	85	(35.7)	129	(35.5)
No hearing loss	29	(23.2)	72	(30.3)	101	(27.8)

Prolonged QTcF interval						
Missing	34	(27.2)	83	(34.9)	117	(32.2)
Normal QTcF	90	(72.0)	148	(62.2)	238	(65.6)
Prolonged QTcF (>450 ms)	1	(0.8)	7	(2.9)	8	(2.2)

Abbreviations: standard deviation (SD), delamanid (DLM), bedaquiline (BDQ), hertz (Hz), decibel (Db), QT interval with Fridericia's correction (QTcF)

Figure 2.1 Flowchart of patients eligible for the comparative effectiveness of delamanid analysis in the endTB observational cohort

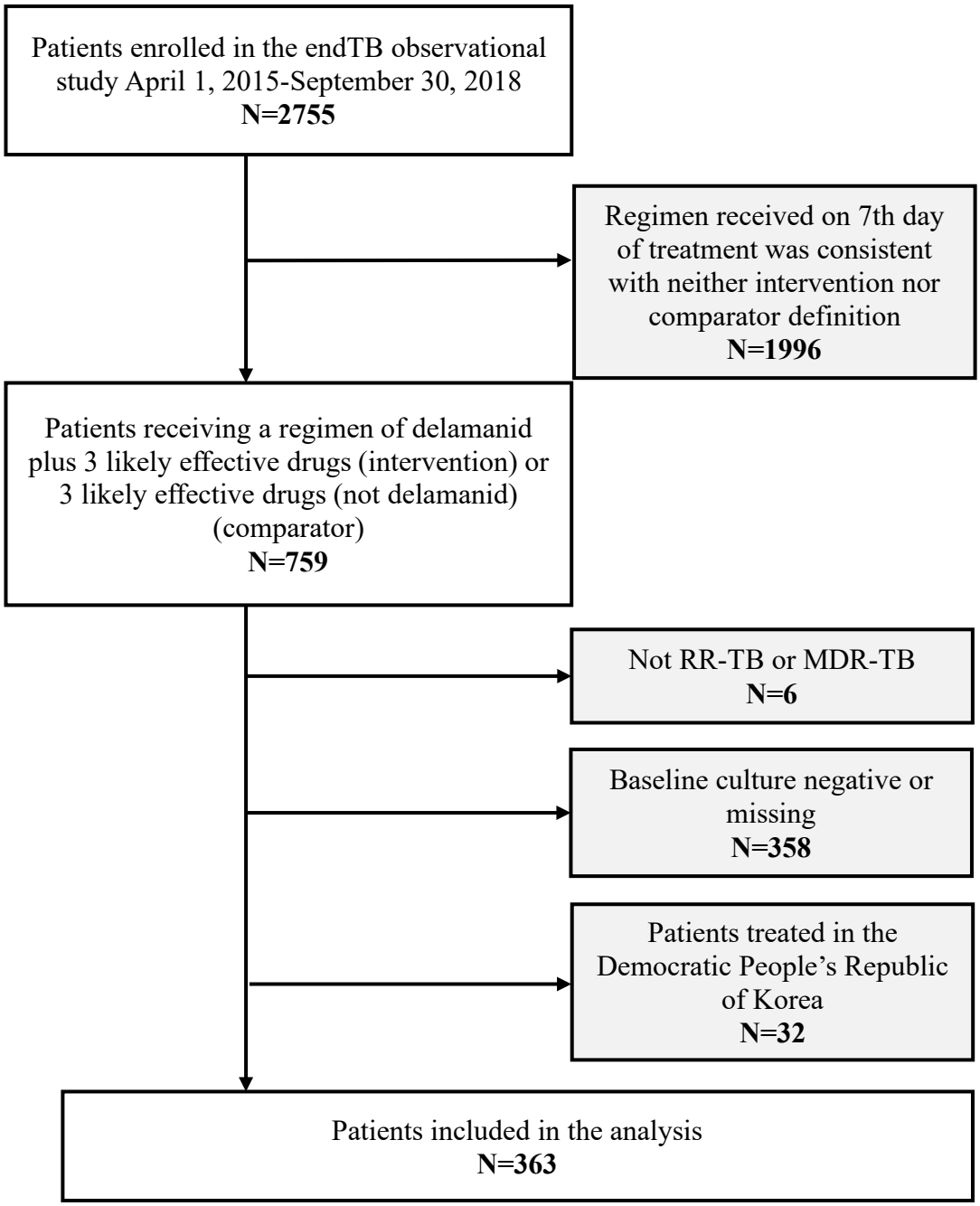


Table 2.2 Baseline treatment regimens of participants receiving three drugs likely to be effective +/- delamanid, endTB observational cohort (N=363)

Baseline regimen	Delamanid containing[a] (N=125)		Delamanid-free (N=238)	
	n	(%)	n	(%)
3 Group A drugs[b]				
Lfx/Mfx, Bdq, Lzd		-	24	(10.1)
2 Group A drugs[b]				
Bdq, Lzd, Cfz	49	(39.2)	142	(59.7)
Lfx/Mfx, Lzd, Cfz	12	(9.6)		-
Bdq, Lzd, ImCil/M	1	(0.8)	20	(8.4)
Bdq, Lzd, Csn/Trd	1	(0.8)	16	(6.7)
Lfx/Mfx, Bdq, Cfz	1	(0.8)	4	(1.7)
Bdq, Lzd, Km/Cm		-	4	(1.7)
Bdq, Lzd, PASms/PAS		-	4	(1.7)
Lfx/Mfx, Lzd, ImCil/M	2	(1.6)		-
Lfx/Mfx, Lzd, Csn/Trd	2	(1.6)		-
Lfx/Mfx, Lzd, Eto/Pto	2	(1.6)		-
Lfx/Mfx, Bdq, Csn/Trd		-	3	(1.3)
Bdq, Lzd, Eto/Pto	1	(0.8)	1	(0.4)
Lfx/Mfx, Bdq, Km/Cm	1	(0.8)		-
Lfx/Mfx, Lzd, Km/Cm	1	(0.8)		-
Bdq, Lzd, E		-	1	(0.4)
Lfx/Mfx, Bdq, E		-	1	(0.4)
Lfx/Mfx, Bdq, PASms/PAS		-	1	(0.4)
1 Group A drug[b]				
Lzd, Cfz, ImCil/M	26	(20.8)		-

Lfx/Mfx, Csn/Trd, Eto/Pto	7	(5.6)		-
Bdq, Cfz, ImCil/M	3	(2.4)	5	(2.1)
Bdq, Cfz, Km/Cm		-	5	(2.1)
Lzd, Cfz, Km/Cm	2	(1.6)		-
Lzd, Cfz, PASms/PAS	2	(1.6)		-
Lzd, Cfz, S/A	2	(1.6)		-
Lfx/Mfx, Cfz, Eto/Pto		-	2	(0.8)
Bdq, Cfz, S/A	1	(0.8)		-
Bdq, Csn/Trd, PASms/PAS	1	(0.8)		-
Lfx/Mfx, Cfz, ImCil/M	1	(0.8)		-
Lfx/Mfx, Eto/Pto, PASms/PAS	1	(0.8)		-
Lzd, Cfz, E	1	(0.8)		-
Lzd, Cfz, Z	1	(0.8)		-
Lzd, ImCil/M, Km/Cm	1	(0.8)		-
Lzd, Csn/Trd, PASms/PAS	1	(0.8)		-
Bdq, Cfz, Eto/Pto		-	1	(0.4)
Bdq, Cfz, PASms/PAS		-	1	(0.4)
Bdq, ImCil/M, Eto/Pto		-	1	(0.4)
Bdq, Csn/Trd, Eto/Pto		-	1	(0.4)
H, Bdq, ImCil/M		-	1	(0.4)
No Group A drugs[b]	1	(0.8)		-
Cfz, Csn/Trd, Eto/Pto	1	(0.8)		-

Abbreviations: bedaquiline (Bdq), clofazimine (Cfz), cycloserine or terizidone (Csn/Tzd), ethambutol (E), ethionamide or prothionamide (Eto/Pto), imipenem-cilastatin-amoxicillin clavulanate or meropenem (ImCil/M), levofloxacin or moxifloxacin (Lfx/Mfx), linezolid (Lzd), p-aminosalicylic acid (PAS), pyrazinamide (Z), streptomycin or amikacin (S/A)

[a] All participants on a delamanid-containing regimen also received delamanid (not shown in regimen list)

[b] Group A drugs include bedaquiline (Bdq), linezolid (Lzd), and levofloxacin or moxifloxacin (Lfx/Mfx)

Figure 2.2 Flowchart of participants censored due to changing treatment strategies in the endTB observational cohort

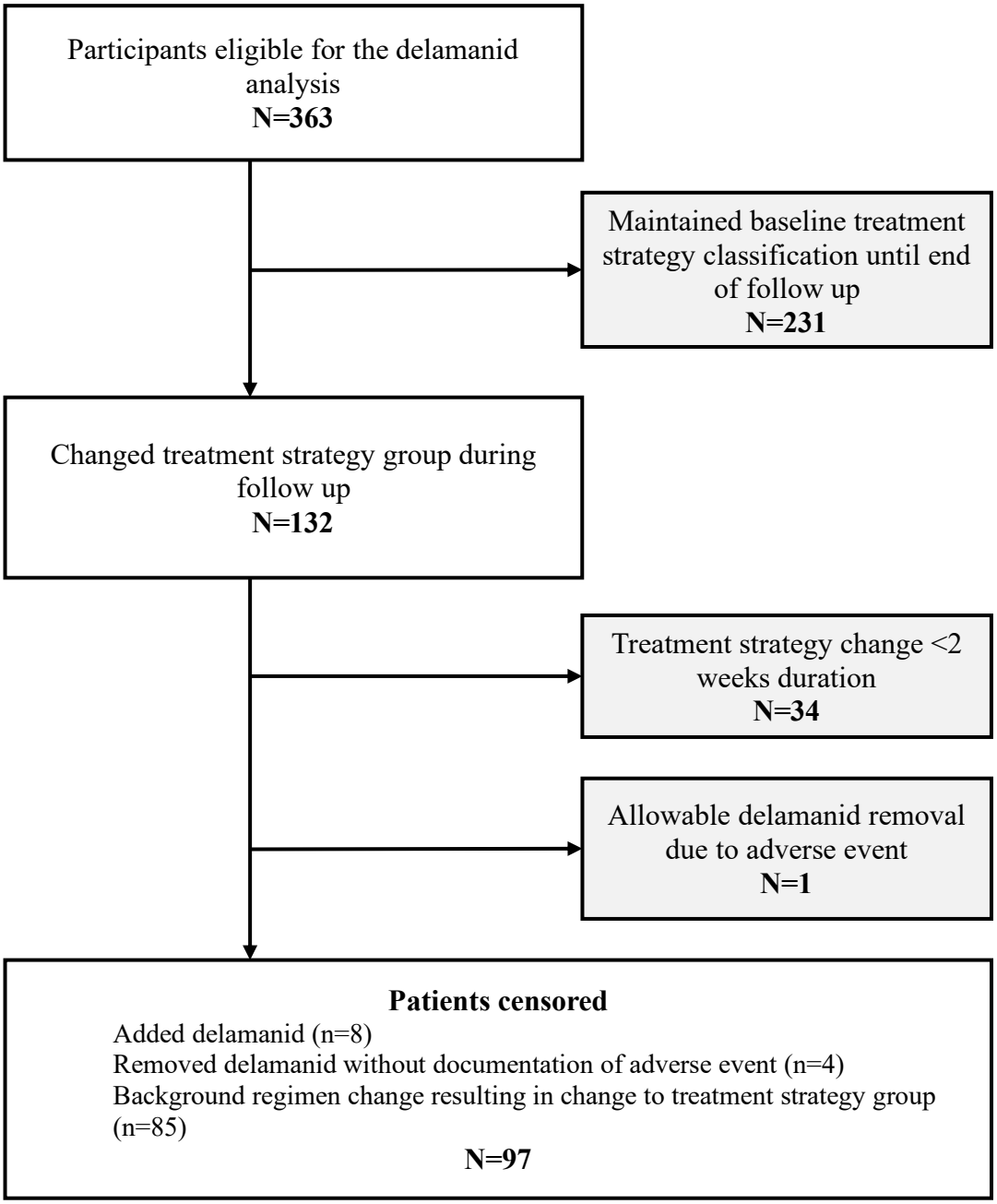


Table 2.3 Effect of adding delamanid to a regimen composed of three drugs likely to be effective, endTB observational cohort (N=363)

Analysis	N	Two-month culture conversion		Six-month culture conversion	
		RR (95% CI)	RD (95% CI)	RR (95% CI)	RD (95% CI)
Crude	363	0.89 (0.72, 1.09)	-0.06 (-0.17, 0.05)	0.91 (0.82, 1.00)	-0.08 (-0.16, -0.00)
IP censoring weighted, unstabilized	349[a]	0.94 (0.67, 1.41)	-0.04 (-0.20, 0.20)	0.93 (0.78, 1.05)	-0.06 (-0.21, 0.05)
IP censoring weighted, stabilized	349[a]	0.94 (0.71, 1.31)	-0.04 (-0.18, 0.16)	0.93 (0.80, 1.05)	-0.06 (-0.19, 0.05)
IP censoring weighted, stabilized, among patients on BDQ	288[b]	0.97 (0.66, 1.42)	-0.02 (-0.20, 0.20)	1.02 (0.94, 1.11)	0.02 (-0.06, 0.10)

Abbreviations: Inverse probability (IP), risk ratio (RR), risk difference (RD), bedaquiline (BDQ)

[a] N=10 participants excluded for missing time-varying data, N=2 participants excluded for missing baseline and time-varying data, N=2 participants excluded for missing baseline data

[b] N=7 participants excluded for missing time-varying data, N=1 participant excluded for missing baseline data

3 SELECTION BIAS IN MULTIDRUG-RESISTANT TUBERCULOSIS COHORT STUDIES USING ASSESSING SPUTUM CULTURE CONVERSION

3.1 Introduction

During treatment of TB, particularly DR-TB, sputum samples are routinely collected and inoculated in culture medium and monitored for growth of *M.tb*. Conversion of sputum culture from positive to negative for *M.tb* is the most important indicator of treatment response.(9,60) Sputum culture conversion is also often used as a proxy for final treatment outcome (e.g. cure, treatment failure, or relapse) in RCTs and observational cohort studies alike.(39,40) While RCTs use standard protocols with sampling at established, frequent intervals, observational cohorts of TB patients are followed under routine conditions. Although the WHO recommends monthly culture monitoring,(61) in this context, patient encounters and sample collection may occur less frequently, and limited laboratory services or reagent stock-outs can result in sparse or inconsistent culture data.

An initial positive sputum culture is a pre-requisite to observe conversion to negative sputum culture. Only patients with positive sputum cultures at a pre-defined "baseline" timepoint (generally expected to be before treatment initiation) are included in such analyses; that is, patients with a negative baseline culture or a missing baseline culture result are excluded. Under routine conditions, patients might not have a documented positive culture immediately before treatment initiation. Therefore, investigators may

define an interval before the day of treatment initiation that constitutes a baseline culture. In some cases, investigators may even extend this interval past treatment initiation in order to include patients who lacked a positive pre-treatment culture but had a positive culture after treatment initiation. While this latter approach can improve precision by increasing sample size, a post-treatment initiation extension could also introduce selection bias. This is because patients added to the analysis cohort will include those without a positive pre-treatment culture who survived or were retained long enough to have a recorded positive baseline culture in the allowable post-treatment interval, but exclude those with a missing pre-treatment culture who die or are LTFU during this interval, events that are often defined as non-conversions. Here, we refer to this bias mechanism as selection bias. Inclusion in the study requires a subset of patients (i.e. patients missing a pre-treatment culture) to survive or be retained in the study long enough to make it past a selection process (i.e. having a culture in the post-treatment initiation interval).(62) The ramifications of selection bias have been assessed at length in the epidemiologic literature, both in concept and applied in data from substantive areas such as perinatal epidemiology,(63–66) trauma,(67) cancer,(68) and HIV.(69)

Using simulations and observational data from a cohort of patients treated for DR-TB, we investigate the potential for bias in reporting the absolute proportion of a cohort with culture conversion when extending the baseline culture definition interval past treatment initiation.

3.2 **Methods**

3.2.1 Absolute proportion of patients with sputum-culture conversion

3.2.1.1 Quantifying bias using simulated data

To investigate selection bias due to extending the baseline culture definition interval past treatment initiation, we first simulated a hypothetical cohort without missing data and then introduced missingness. Conducting bias analyses in real-world data is limited by the range of values observed in the dataset. By simulating data, we can generate the full range of potential values for each parameter (e.g. the proportion of patients culture positive, the proportion of patients with a missing pre-treatment culture) and better understand how parameters interact.(70,71) The outcome of interest was culture conversion, defined as two, consecutive negative cultures at least 15 days apart. Death or LTFU before conversion were treated as non-conversion events. To simulate a hypothetical cohort, we specified the parameters below using the values listed in Table 3.1:

- *Culture positive*$_{truth}$ (P_t), represents the proportion of patients who would have been observed to be culture positive at the time of treatment initiation, had they had a sputum culture result. The proportion of culture negative patients is therefore calculated by 1-P_t.
- *Culture missing*$_{observed}$ (m), represents the proportion of patients missing a pre-treatment culture

- *Culture positive*$_{truth}$ | *Culture missing*$_{observed}$ ($P_t \mid m$), represents the proportion of patients with a missing pre-treatment culture, who would have been observed to have a positive culture at treatment initiation, had they had a culture
- *Converted* | *Culture positive*$_{truth}$ ($C \mid P_t$), represents the proportion of patients with conversion among patients who would have been observed to have a positive culture at the time of treatment initiation, had they had a sputum culture result

The proportion with culture conversion is calculated by dividing the number of patients observed to have converted by the number of patients observed with a positive baseline culture. We calculate the observed proportion with culture conversion (*Converted* | *Culture positive*$_{observed}$ ($C \mid P_O$)) from the aforementioned parameters as follows: $\frac{P_t \times C \mid P_t}{P_t - (m \times P_t \mid m)}$. This formula effectively includes all conversion events in the numerator and subtracts from the denominator patients missing a culture who would have been positive, had they had a culture. We hypothesized that early deaths and LTFUs occurring during the post-treatment allowable interval would drive the differences between true and observed conversion frequencies, and therefore that conversion in patients missing a pre-treatment culture occurred at an equal or lower frequently than among patients observed to have a pre-treatment culture. This assumption effectively limits $C \mid P_O$ to values greater than or equal to $C \mid P_t$.

We report the difference between $C \mid P_t$ and $C \mid P_O$ and the minimum and maximum proportion of the cohort observed to have converted and percentage point discrepancy between the two figures. We excluded combinations of values producing results that

exceeded the 0.00 to 1.00 bounds of a proportion.

3.2.1.2 Quantifying maximum bias in the endTB observational cohort

In order to determine the extent to which bias might have impacted a real cohort of DR-TB patients, we used data from the endTB observational cohort (ClinicalTrials.gov record NCT02754765). The endTB observational cohort is a prospective cohort of patients with DR-TB treated with bedaquiline and/or delamanid. Patients were eligible for inclusion if they initiated an endTB treatment regimen between 04/01/2015 and 11/16/2018. Data collection followed a standardized common protocol, which has been previously described.(30)

We compared a baseline culture definition with an allowable interval of 90 days before (-90) and 0 days after (+0) treatment initiation to definitions that extended the allowable interval to 30, 60 and 90 days past treatment initiation (-90/+30 days, -90/+60 days, -90/+90 days). For the latter definitions, patients who had any positive culture(s) during the specified interval were considered to have a positive baseline culture; those who had consistently negative culture(s) during the specified interval were considered to have a negative baseline culture. Patients without any cultures during the specified interval were classified as having a missing baseline culture.

We used an interim endpoint of six-month culture conversion. While definitions of six-month culture conversion are not standardized and vary considerably in the literature,(72) our definition requires two consecutive negative cultures collected at least 15 days apart, the first occurring up to 180 days after treatment initiation and the second up to 210 days after treatment initiation. For baseline culture definitions extending past treatment

initiation (-90/+30, -90/+60, -90/+90), negative cultures fulfilled the culture conversion definition only if they were performed on sputum samples that were collected after the first positive culture had been established. Participants who died or were lost to follow-up before sputum culture conversion were considered not to have converted because these events are considered unfavorable final TB treatment outcomes.(73) Death was defined as death due to any cause and LTFU was defined as treatment interruption for two or more consecutive months.

For each of the four baseline culture intervals, we quantified the observed proportion of the cohort with sputum-culture conversion ($C \mid P_O$) as follows: $\frac{N\ Converted}{N\ Culture\ positive_{observed}}$. When the interval is extended past treatment initiation, we report the additional number of participants added to the analysis and the number of participants missing a baseline culture who died or were LTFU during the interval. In order to investigate the upper bound of the magnitude of bias, we calculated the proportion converted "assuming maximum bias" (Converted $\mid$ Culture positive$_{\text{maximum bias,}}$ abbreviated $C \mid P_{mb}$) as follows: $\frac{N\ Converted}{N\ Culture\ positive_{observed}\ +\ N\ died\ or\ LTFU \mid Culture\ missing_{observed}}$.

This equation reflects maximum bias in that it presumes that 100% of patients with a missing pre-treatment culture who died or were LTFU during the post-treatment initiation culture interval would have been observed to be culture-positive, had they received a culture. As this percentage decreases from 100%, the expected magnitude of bias decreases.

3.2.2 Relative proportion of patients with sputum-culture conversion

Using data from the endTB observational cohort, we assessed the impact of extending the baseline culture interval past treatment initiation on the crude relative risk of common six-month culture conversion predictors. Risks were calculated with the same formulas as $C \mid P_O$ and $C \mid P_{mb}$. To assess bias we consider log risk ratios from models using definitions extending the baseline culture interval past treatment initiation (i.e. -90/+30, -90/+60, -90/+90 days) to be represented by $\beta_{-90/+\geq 1}$. The log risk ratio using a definition that does not extend the baseline culture interval past treatment initiation (i.e.-90/+0) is represented by $\beta_{-90/+0}$. The percentage of bias in $\beta_{-90/+\geq 1}$ due to extending the culture collection interval past treatment initiation is thus represented by $\frac{\exp(\beta_{-90/+\geq 1}) - \exp(\beta_{-90/+0})}{\exp(\beta_{-90/+0})} \times 100$. To assess precision, we calculate the mean squared error for each estimate as follows: $(\beta_{-90/+\geq 1} - \beta_{-90/+0})^2 + (\mathrm{SE}(\beta_{-90/+\geq 1})^2$.

We considered the following dichotomous predictors: diabetes mellitus or glucose intolerance, poorly controlled diabetes (HbA1c >8.0%), HIV infection, CD4 cell count <200 cells/mL, Hepatitis B virus infection, Hepatitis C virus infection, prior TB treatment with second-line drugs, bilateral disease, cavitary disease, and BMI <18.5.

3.3 **Results**

3.3.1 Absolute proportion of patients with sputum-culture conversion

3.3.1.1 *Quantifying bias using simulated data*

Combinations of the values listed in Table 3.1 resulted in 420 potential scenarios; 66

exceeded the 0.00 to 1.00 bounds for a proportion and were excluded from the analysis, leaving 354 results. Figure 3.1 provides a guided interpretation of simulation study results presented in Figures 3.2-3.4 and annotates two of the most influential drivers of bias. In brief, the colored horizontal line represents the true proportion of the cohort with culture conversion ($C \mid P_t$). The shaded region indicates the potential minimum and maximum bias as a function of the proportion of patients who died or were LTFU among those missing a pre-treatment at the value of m on the x-axis. Thus, the upper bound of the shaded region (indicated by ■ in Figure 3.1) represents the proportion observed if all patients missing a pre-treatment culture had died or been LTFU (i.e. did not convert). The lower bound of the shaded region (indicated by ● in Figure 3.1) reflects the point at which conversion rates in patients missing a pre-treatment culture and patients observed to have a pre-treatment culture were the same. This value is equal to $C \mid P_t - 1$. The two most influential drivers of bias, as shown by the width of the shaded region, are the proportion of patients missing a pre-treatment culture (x-axis) and the proportion of these patients who died or were LTFU (vertical point within shaded region).

Four key patterns emerged from the simulation. First, the potential for bias increases as the proportion missing a pre-treatment culture (m) increases, as shown by the larger width of shaded regions at higher values of the x-axis in Figures 3.2-3.4. Second, the potential for bias is limited in cohorts with high rates of conversion, as shown by the leveling off of the shaded region's upper bound at 100% where $C \mid P_t$=90% and m=~15% in Figure 3.2. This is because the upper bound of the shaded region reflects a scenario in which all patients missing a pre-treatment culture die or are LTFU, both of which are

considered non-conversion events, during the allowable interval. Third, holding other parameters fixed, as the proportion of patients in a cohort who are truly culture positive at treatment initiation (P_t) increases from 60% to 90%, the potential magnitude of maximum bias decreases, as shown by the decreasing shaded regions' widths across the panel in Figure 3.3. This is because the exclusion of the same number of patients from a cohort with a smaller denominator (e.g. 60% of cohort is culture positive) is more influential on the observed proportion with culture conversion than in a cohort with a larger denominator (e.g. 90% of cohort is culture positive). Lastly, the magnitude of maximum bias is dependent on the proportion of patients who are truly culture positive among those missing a pre-treatment culture ($P_t \mid m$) (Figure 3.4). If all patients missing a pre-treatment culture are truly culture negative, these patients would have been excluded from analyses of culture-based endpoints and no bias will be introduced. Conversely, if all patients missing a pre-treatment culture are truly culture positive, these patients should be included and the magnitude of bias will depend on the amount of missingness, early death and LTFU rates in this subset of the cohort, and to a lesser extent, the proportion of patients truly culture positive at treatment initiation.

3.3.1.2 Quantifying maximum bias in the endTB observational cohort data

Between April 6, 2015 and November 16, 2018, 2790 participants initiated a regimen and consented to participation in the endTB observational study. Of these, 1769 participants had a positive pre-treatment culture within 90 days before treatment initiation (-90/+0) (Table 3.2) and a six-month culture conversion outcome (Table 3.3). Assuming a baseline culture definition that does not extend past treatment initiation (-90/+0), 86%

(N=1518/1769) of the cohort achieved culture conversion by 6 months. The observed proportion with culture conversion did not change from 86% when the baseline culture definition was extended to 30, 60, or 90 days past treatment initiation.

Seventeen participants without a culture died or were LTFU in the 30-day period after treatment initiation, the majority of which were in Lesotho, a high HIV burden country. Assuming these participants would have been culture positive if they had had a culture (i.e. maximum bias), the proportion with culture conversion would be one percentage point lower (85%, N=1614/1900) than the observed proportion with culture conversion **(**Table 3.3 and Figure 3.5 -90/+30 days). The potential magnitude of maximum bias differed by site. Deaths or LTFU events in the first 30 days of treatment occurred among participants missing a pre-treatment culture in 6/17 (35%) sites (Table 3.3). We report a one to five percentage point discrepancy between the observed proportion with culture conversion and proportion with culture conversion assuming maximum bias (Figure 3.5, -90/+30 days) in these 6 sites.

In the 60- and 90-day periods after treatment initiation, death and LTFU events among participants missing a culture increased to 24 and 26, respectively (Table 3.3). Among all participants, the proportion with culture conversion assuming maximum bias was two percentage points lower (84%) than the observed, based on a definition that extended the baseline culture definition 60 or 90 days after treatment initiation. A death or LTFU event among participants missing a culture occurred in 8/17 (47%) sites (Table 3.3), resulting in a reported proportion with culture conversion of up to a 5 percentage point overestimate (Figure 3.5, -90/+60 and -90/+90 days). One site (Kenya) had a 10

percentage point discrepancy using a -90/+90 definition (Figure 3.5, -90/+90 days); however, this site had an extremely small sample size (N=4).

3.3.2 Relative proportion of patients with sputum-culture conversion

3.3.2.1 endTB observational cohort data

We show the crude log relative risk of predictors on sputum culture conversion in the endTB observational cohort when extending the baseline culture definition 30 (Figure 3.6), 60 (Figure 3.7) and 90 days (Figure 3.8) after treatment initiation. Using a definition that does not extend the sputum collection interval past treatment initiation as a reference, the largest discrepancies were observed for HIV infection and CD4 cell count <200 cells/mL. Extending the sputum collection interval by 30 days past treatment initiation resulted in a 2.3% change in the relative risk for HIV (0.83 at -90/+0 days vs 0.85 at -90/+30 days); assuming maximum bias drew the point estimate back towards the reference estimate (Table 3.4). Similar results were observed in the -90/+60 (Table 3.5) and -90/+90 days (Table 3.6) analyses. When accounting for death and LTFU after treatment initiation, bias in CD4 cell count <200 cells/mL point estimates exceeded 5% in the -90/+60 (Table 3.5) and -90/+90 days (Table 3.6) analyses. MSE differences across estimates were negligible.

3.4 **Discussion**

Using simulated data and in a real-world cohort study, we investigated the potential for selection bias when extending baseline culture definitions to include a period after

treatment initiation. Our findings have important implications for investigators reporting on basic epidemiologic parameters of DR-TB treatment cohorts, including the proportion with culture conversion and the relative risk of common predictors of culture conversion. In the present analyses, we identified the most influential factors that increased bias were the proportion of the cohort with a missing pre-treatment culture and the occurrence of death and LTFU in this group.

Simulation studies are a valuable tool for exploring complex theoretical biases in epidemiology, which can be especially useful in contexts where a "true" dataset will never be realized, such as in the case of selection bias. However, simulations are only useful insofar they are encoded with values that can play out in real life.(70) In our simulation, we present the entire spectrum of potential bias in the reported proportion with culture conversion. In reality, conversion overestimates are likely much less than the maximum presented in simulations, as shown by the upper bound of the shaded region, because it is improbable everyone missing a pre-treatment culture dies or is LTFU in the post-treatment initiation interval. These smaller estimated changes would suggest that a missing pre-treatment culture is not necessarily due to early death or LTFU. A large overestimate (e.g. 20-30 percentage points) could occur if a large percentage (20–30%) of culture-positive patients were missing a pre-treatment culture and the majority of these patients died or were LTFU in the post-treatment initiation interval. However, high rates of early death are uncommon in today's cohorts given advances in treatment that have drastically reduced mortality.(74,75) However, the same cannot be said for historical cohorts. High early death rates were common among patients with advanced drug-

resistance and HIV co-infection, such as in South Africa.(76) In fact, among patients initiating DR-TB treatment in South Africa between 2012 and 2014, 10% of the cohort died within the first 12 weeks of treatment and a missing or contaminated baseline culture was the strongest predictor of mortality (hazard ratio: 3.78, 95% CI: 2.90-3.99).(77) While this study did not assess culture conversion, it does serve as an example of a cohort for which extending the allowable baseline interval past treatment initiation could introduce bias.

We identified notable heterogeneity across sites in analyses assessing maximum bias of the proportion with culture conversion in the endTB cohort. These site-specific differences provide some insight into how this mechanism of selection bias can play out across settings with vastly different prevalence of comorbidities. For example, high rates of early death and LTFU occurred in Lesotho, potentially due to a higher prevalence of HIV and advanced drug resistance. And, in Kenya, even one or two early death or LTFU events drastically biased the proportion with conversion due to the site's small sample size. Many reports of MDR-TB clinical cohorts are confined to small groups of patients treated in the same geographic settings with similar comorbidities. Results from the simulation also reinforce that low rates of conversion and low proportions of patients with a positive pre-treatment culture increase the potential for bias, two factors that are likely common in cohorts comprised of people living with HIV. Investigators reporting on patients with comorbidities or other factors known to predict early death or LTFU should report whether these early events manifest in their cohort and the pre-treatment culture status of such patients.

Two circumstances must be at play in order for relative measures to differ in the presence of selection bias from baseline culture definitions extending past treatment initiation. First, a sizable proportion of the cohort must be missing a pre-treatment culture and die or be LTFU in the allowable post-treatment initiation collection interval. Second, the predictor must be highly associated with both missingness and early death or LTFU. That is, the majority of patients missing a pre-treatment initiation culture who die or are LTFU must be concentrated within one level of the predictor. If these patients are equally distributed between levels of the predictor, no bias will be introduced into the relative risk estimate. In the endTB cohort, we found little meaningful change that would affect study interpretation of relative measures because, although 12% of the cohort was missing a pre-treatment culture, missingness was not highly associated with early death and LTFU. The largest shifts in point estimates across baseline culture definitions were observed for HIV and CD4 count <200 cells/mm^3. This is a logical finding, given HIV coinfection and the immune status of people living with HIV are highly predictive of mortality.(78,79) The magnitude and direction of selection bias for relative measures has been studied extensively, both through simulation using directed acyclic graphs and applied within a variety of subject areas.(62,63,80–82) Concordant with our findings, these studies also reveal that the magnitude of missingness alone is insufficient to infer whether there is bias. It also depends on the proportion of those with the outcome among those with missing data.

Investigators may be tempted to extend the allowable baseline culture collection interval past treatment initiation in order to capture more patients, thereby improving study

precision. However, in our analysis we observed no improvements in precision despite the addition of 6% of the original cohort to analyses extending the culture collection interval 30 days past treatment initiation. A precision advantage may be observed when adding a relatively large proportion of the total cohort to the analysis, especially in a smaller cohort(83). However, the addition of a large proportion of the sample may increase the probability of bias, as demonstrated by our simulation findings in which overestimates of conversion were larger in cohorts with more missingness of pre-treatment cultures.

Several steps can be taken to prevent and assess the potential for this bias. In the analytic phase, investigators could implement baseline culture definitions that do not extend the allowable collection interval after treatment initiation. While this definition may exclude some patients who had a culture shortly after treatment initiation, it also eliminates the potential for bias. If investigators extend the culture collection interval definition past treatment initiation, the investigator can simply check for death or LTFU events that occurred in the post-treatment initiation interval among those without a pre-treatment culture. If these events occurred, an investigator should consider how these events were distributed across the predictor or treatment of interest and to what extent excluding these patients will introduce selection bias into the final estimate. During the data collection phase, investigators can avoid the potential for bias altogether by making dedicated efforts to collect sputum specimens before or on the day of treatment initiation. But, securing complete pre-treatment culture data is undoubtedly difficult in the context of observational DR-TB treatment cohorts. Patients may have difficulty producing sputum

and operational challenges to obtaining a sputum specimen, such as few laboratories equipped to conduct culture testing, impose barriers having pre-treatment culture results. Additionally, priority has been placed on decentralized capacity for rapid molecular tests (e.g. Xpert® MTB/RIF (Cepheid, Sunnyvale, USA)) to diagnose pulmonary TB and detect rifampin resistance, which was recommended by the WHO in 2020.(84) This may reduce availability of (pre-treatment) sputum culture results.

We present potential bias estimates for the proportion with culture conversion disaggregated by each of the 17 enrollment sites in order to highlight heterogeneity that may arise from settings of different patient characteristics (e.g. comorbidities) and early treatment outcomes. Some countries had relatively small sample sizes. In these small sub-cohorts, even a few patients with a missing pre-treatment culture who die or are LTFU early in treatment can impose substantial bias in the proportion with culture conversion. Small sample sizes at these sites are not necessarily a study limitation. In fact, they reflect the size of cohorts routinely reported on the MDR-TB literature: approximately 8% of MDR-TB cohort studies in the last five years reported on less than 25 patients and 31% reported on less than 100.(72) Second, in the endTB absolute proportion analysis, we calculate the proportion with culture conversion assuming maximum bias by adding to the denominator 17 patients who were missing a pre-treatment culture and died or were LTFU during the 30-day interval after treatment initiation. An additional 154 patients missing a pre-treatment culture were retained during this same period. We did not pursue in-depth analyses to assess how the exclusion of retained patients affected the observed proportion. Rather, we assume patients with a missing pre-treatment culture convert at an

equal or lower frequency than those with a pre-treatment culture because our aim was to quantify the upper bound of bias (i.e. overestimates) in the proportion with culture conversion.

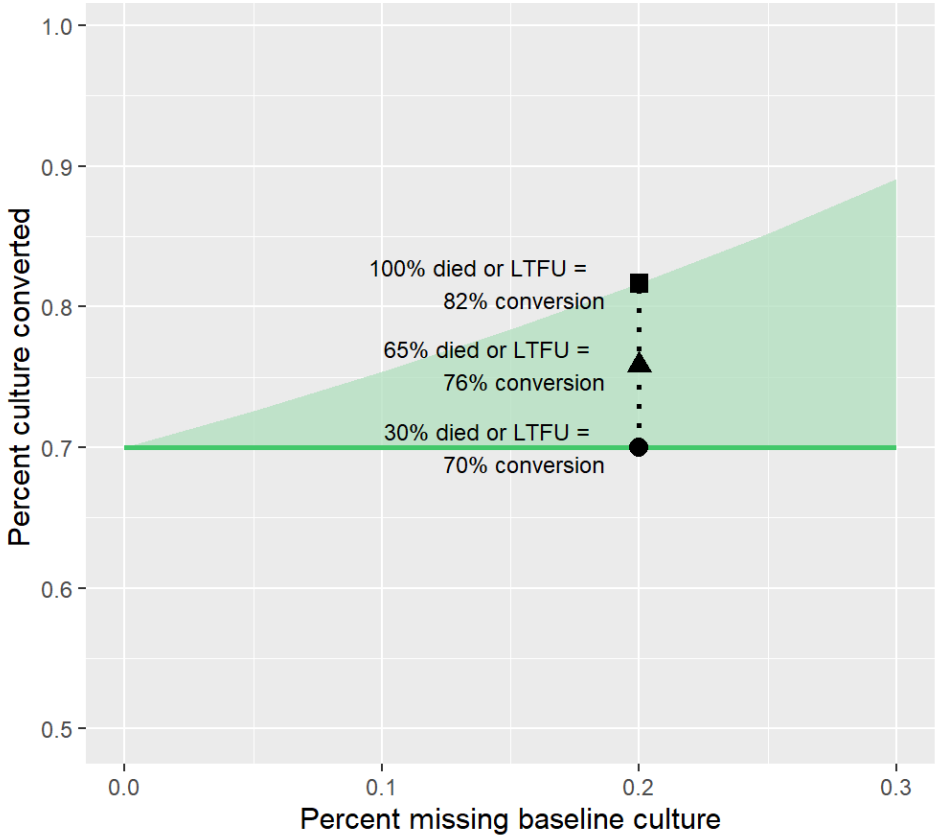
The implications of our study findings underscore the need to scrutinize whether bias is introduced when determining who is included and excluded from analyses with culture-based endpoints. Avoiding extension of the baseline culture collection interval past treatment initiation will eliminate the potential for bias. When this definition is extended past treatment initiation, the decision to do so should be clearly reported and early death and LTFU events among excluded patients should be enumerated. Taking these steps will improve transparency and comparability of study findings across cohorts, thereby improving the evidence based used to inform the treatment of the estimated 500,000 new DR-TB patients each year.

3.5 Tables and Figures

Table 3.1 Parameters and values, simulation study of bias due to early death and loss-to-follow up events occurring during a hypothetical post-treatment initiation sputum collection interval among participants missing a pre-treatment sputum culture

Parameter	Description	Simulated values
Culture positive$_{truth}$ (P_t)	Proportion of patients who would have been observed to be culture positive at the time of treatment initiation, had they had a sputum culture result	60% to 90% by 10%
Culture missing$_{observed,}$ (m)	Proportion of patients observed missing a pre-treatment culture	0% to 30% by 5%
Culture positive$_{truth}$ | *Culture missing*$_{observed}$ (P_t | m)	Proportion of patients with a missing pre-treatment culture who would have been observed to have a positive culture at treatment initiation, had they had a culture	0% to 100% by 25%
Converted | *Culture positive*$_{truth}$ (C | P_t)	Proportion of patients with conversion among patients who would have been observed to have a positive culture at the time of treatment initiation, had they had a sputum culture result	50% to 90% by 20%

Figure 3.1 Culture conversion and early death and loss-to-follow up events among participants missing a pre-treatment culture, simulation study of cohort where Pt=70%, Pt | m=50%, and C | Pt=70%

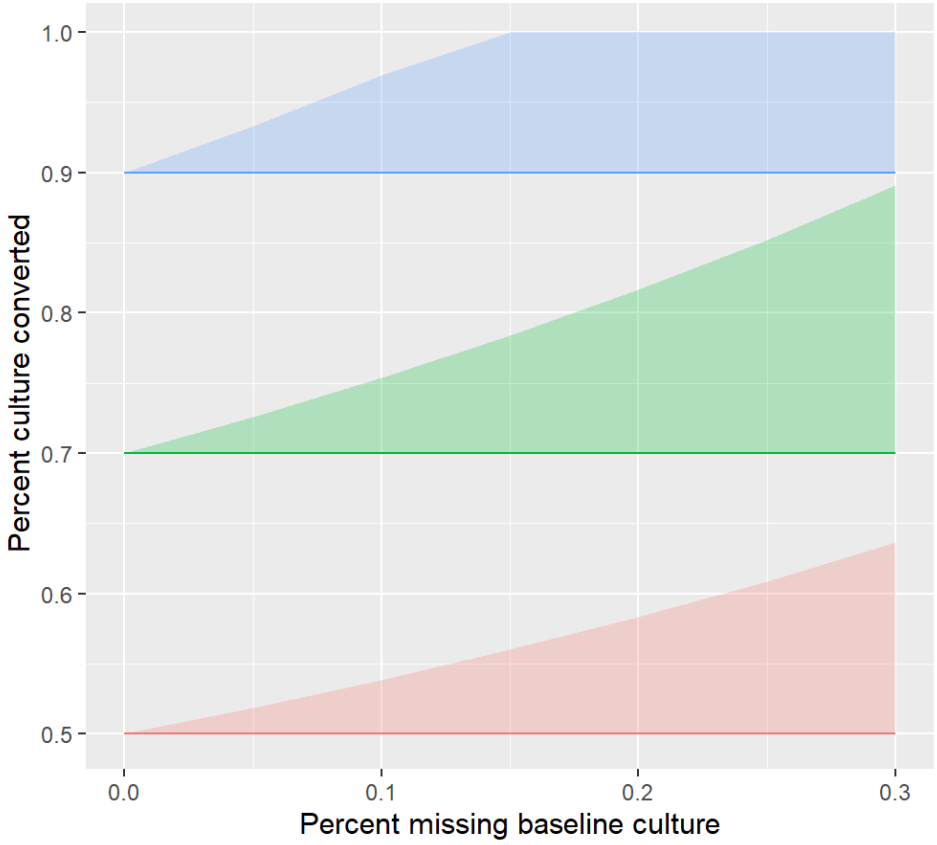


Legend: Figure represents a simulated cohort of patients in which: 1) 70% are truly culture positive, 2) Of patients missing a culture, 50% are truly culture positive and 50% are truly culture negative, and 3) among truly culture-positive patients, 70% achieved culture conversion. If 20% (x-axis=0.20) of patients were missing their pre-treatment culture and 100% of these patients died or were LTFU during the hypothetical post-treatment initiation sputum collection interval, the observed proportion with culture conversion would be 82% (■), a 12 percentage point discrepancy. If 65% (i.e. the

halfway point of the shaded region) died or were LTFU, the reported proportion would be 76% (▲), a 6 percentage point discrepancy. If 30% died or were LTFU (i.e. the point at which conversion rates in patients missing a pre-treatment culture and patients observed to have a pre-treatment culture are equal), the reported proportion would be 70% (●), no discrepancy. The shaded region can be interpreted similarly in Figures 2-4.

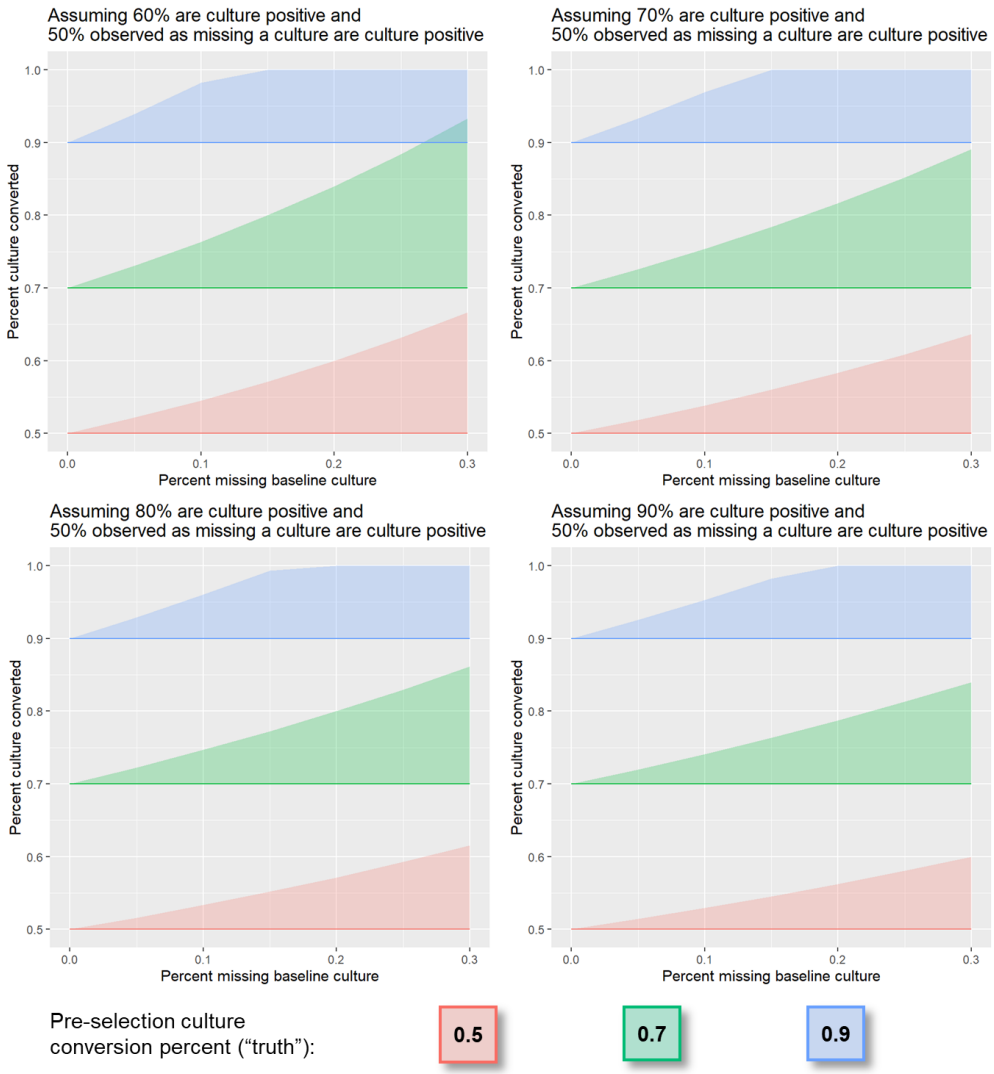
Abbreviations: Loss to follow up (LTFU); Culture positive$_{truth}$ (P_t); Culture positive$_{truth}$ | Culture missing$_{observed}$ (P_t | m); Converted | Culture positive$_{truth}$ (C | P_t)

Figure 3.2 **Culture conversion and early death and loss-to-follow up events among participants missing a pre-treatment culture, simulation study of cohort where Pt=70%, Pt | m=50%, and C | Pt=50%, 70%, or 90%**



Legend: In cohorts with high rates of culture conversion (e.g. 90% conversion in blue), there a modest number of patients can be missing a pre-treatment culture (e.g. 15% at 90% conversion), assuming missing a pre-treatment culture is perfectly correlated with death or LTFU (upper bound of shaded region).

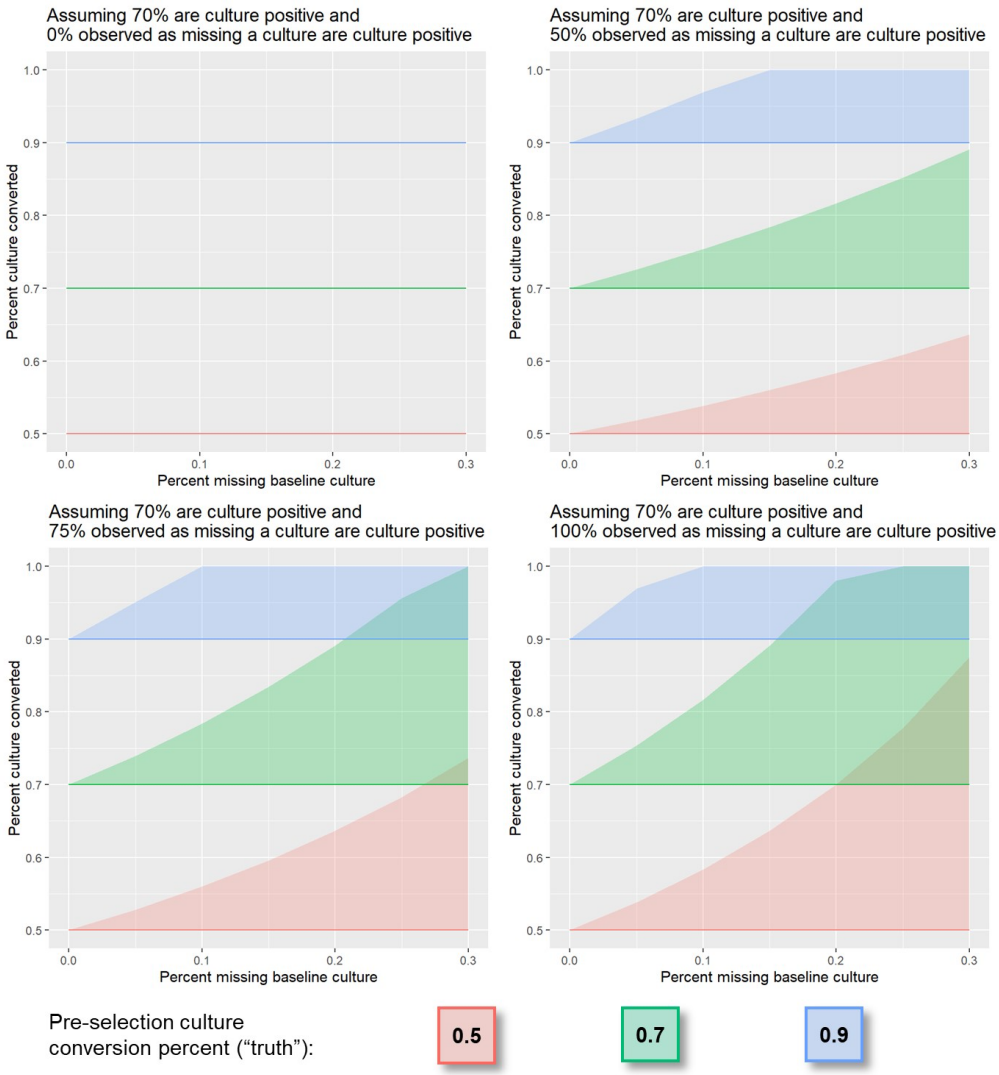
Figure 3.3 Culture conversion and early death and loss-to-follow up events among participants missing a pre-treatment culture, simulation study of cohort where Pt varies from 60% to 90%, Pt | m=50%, and C | Pt=50%, 70%, or 90%



Legend: As the proportion of patients who are culture positive at treatment initiation increases (each panel), the potential magnitude of bias decreases (shared regions become smaller). This is because, assuming the proportion missing a pre-treatment culture who

are culture positive is held constant (here, 50%), the exclusion of the same number of patients from a smaller cohort (i.e. smaller denominator) is more influential on the observed proportion than in a larger cohort (i.e. larger denominator).

Figure 3.4 Culture conversion and early death and loss-to-follow up events among participants missing a pre-treatment culture, simulation study of cohort where Pt=70%, Pt | m varies from 250-100% and C | Pt=50%, 70%, or 90%



Legend:

The maximum magnitude of bias (top of each shaded region) is dependent on the proportion of patients who are culture positive among those missing a pre-treatment

culture (each panel). If no patients (0%) missing a pre-treatment culture are culture positive, then these patients would be excluded from the analysis and no bias will be introduced (top left). Conversely, if all (100%) patients missing a pre-treatment culture are culture positive, then these patients should be included and the most potential for bias is introduced (bottom right).

Table 3.2 Pre-treatment initiation culture status of participants in the endTB observational cohort (N=2790)

Country	-90/+0 days			
	Participants enrolled, N	Positive culture (-90/+0), n (%)	Negative culture (-90/+0), n (%)	Missing culture (m) (-90/+0), n (%)
Armenia	107	86 (0.80)	14 (0.13)	7 (0.07)
Bangladesh	280	187 (0.67)	68 (0.24)	25 (0.09)
Belarus	109	73 (0.67)	3 (0.03)	33 (0.30)
Ethiopia	79	34 (0.43)	24 (0.30)	21 (0.27)
Georgia	291	214 (0.74)	58 (0.20)	19 (0.07)
Haiti	37	24 (0.65)	2 (0.05)	11 (0.30)
Indonesia	72	40 (0.56)	8 (0.11)	24 (0.33)
Kazakhstan	672	418 (0.62)	226 (0.34)	28 (0.04)
Kenya	7	3 (0.43)	2 (0.29)	2 (0.29)
Kyrgyzstan	18	13 (0.72)	1 (0.06)	4 (0.22)
Lesotho	264	127 (0.48)	53 (0.20)	84 (0.32)
Myanmar	50	16 (0.32)	14 (0.28)	20 (0.40)
North Korea	155	80 (0.52)	41 (0.26)	34 (0.22)
Pakistan	302	246 (0.81)	46 (0.15)	10 (0.03)
Peru	266	158 (0.59)	105 (0.39)	3 (0.01)
South Africa	49	26 (0.53)	9 (0.18)	14 (0.29)
Vietnam	32	27 (0.84)	5 (0.16)	0 (0)
Total	2790	1772 (0.64)	679 (0.24)	339 (0.12)

Table 3.3 Sputum culture conversion and early death and loss-to-follow up events among participants missing a sputum culture in the specified interval before (-) and after (+) treatment initiation, endTB observational cohort

Country	-90/+0 days	-90/+30 days		-90/+60 days		-90/+90 days	
	C | P_O[a], n/N (%)	C | P_O[a], n/N (%)	Died/LTFU, 1-30 days | m, N	C | P_O[a], n/N (%)	Died/LTFU, 1-60 days | m, N	C | P_O[a], n/N (%)	Died/LTFU, 1-90 days | m, N
Armenia	56/86 (0.65)	56/89 (0.63)	1	56/89 (0.63)	1	56/89 (0.63)	1
Bangladesh	182/187 (0.97)	189/194 (0.97)	0	192/197 (0.97)	0	193/198 (0.97)	0
Belarus	60/73 (0.82)	74/88 (0.84)	0	82/96 (0.85)	0	83/97 (0.86)	0
Ethiopia	29/34 (0.85)	33/39 (0.85)	0	33/39 (0.85)	0	33/39 (0.85)	0
Georgia	188/214 (0.88)	195/221 (0.88)	0	198/225 (0.88)	0	198/225 (0.88)	0
Haiti	16/24 (0.67)	17/25 (0.68)	0	17/26 (0.65)	0	17/26 (0.65)	0
Indonesia	27/40 (0.68)	33/48 (0.69)	4	33/50 (0.66)	4	33/51 (0.65)	5
Kazakhstan	400/418 (0.96)	414/433 (0.96)	1	419/440 (0.95)	1	421/442 (0.95)	1
Kenya	1/3 (0.33)	2/4 (0.50)	0	2/4 (0.50)	1	2/4 (0.50)	1
Kyrgyzstan	10/13 (0.77)	12/15 (0.80)	0	12/15 (0.80)	0	12/15 (0.80)	0
Lesotho	90/127 (0.71)	108/150 (0.72)	8	111/155 (0.72)	10	111/155 (0.72)	10
Myanmar	14/16 (0.88)	15/17 (0.88)	0	16/18 (0.89)	1	17/19 (0.89)	1

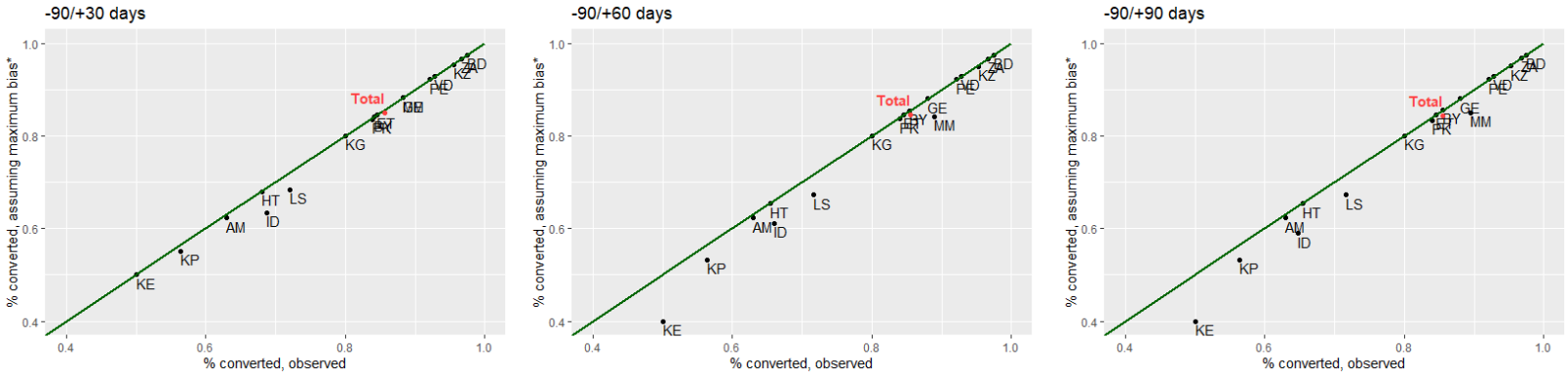
North Korea[b]	42/77 (0.55)	49/87 (0.56)	2	49/87 (0.56)	5	49/87 (0.56)	5
Pakistan	207/246 (0.84)	209/249 (0.84)	1	210/250 (0.84)	1	210/250 (0.84)	2
Peru	146/158 (0.92)	153/166 (0.92)	0	153/166 (0.92)	0	153/166 (0.92)	0
South Africa	25/26 (0.96)	29/30 (0.97)	0	29/30 (0.97)	0	30/31 (0.97)	0
Vietnam	25/27 (0.93)	26/28 (0.93)	0	26/28 (0.93)	0	26/28 (0.93)	0
Total	1518/1769 (0.86)	1614/1883 (0.86)	17	1638/1915 (0.86)	24	1644/1922 (0.86)	26

Abbreviations: Lost to follow up (LTFU), *Culture missing*$_{observed,}$ (m); Converted | Culture positive$_{observed}$ (C | P$_O$)

[a] Observed proportion of the cohort with sputum-culture conversion, Converted | Culture positive$_{observed}$ (C | P$_O$) = $\frac{N\ Converted}{N\ Culture\ positive_{observed}}$

[b] N=3 patients in North Korea do not have a six-month culture outcome and are excluded from the analysis

Figure 3.5 Absolute proportion of sputum culture conversion in the endTB observational cohort, by site and allowable baseline sputum culture collection interval before (-) and after (+) treatment initiation



*Proportion of the cohort with sputum culture conversion, assuming maximum bias ($\%\ Converted \mid Cu +_{\max bias}$) was calculated as follows: $\frac{N\ Converted}{N\ Culture\ positive_{observed} + N\ died\ or\ LTFU \mid Culture\ missing_{observed}}$.

Legend: Sites on the green line indicate no deaths or LTFU events among participants with a missing culture occurred in the specified interval before (-) and after (+) treatment initiation.

Abbreviations: Armenia (AM), Bangladesh (BD), Belarus (BY), Ethiopia (ET), Georgia (GE), Haiti (HT), Indonesia (ID), Kazakhstan (KZ), Kenya (KE), Kyrgyzstan (KG), Lesotho (LS), Myanmar (MM), North Korea (KP), Pakistan (PK), Peru (PE), South Africa (ZA), Vietnam (VD)

Figure 3.6 Crude relative risk of sputum culture conversion by baseline characteristics, comparison extending the interval 30 days past treatment initiation to no extension of interval past treatment initiation

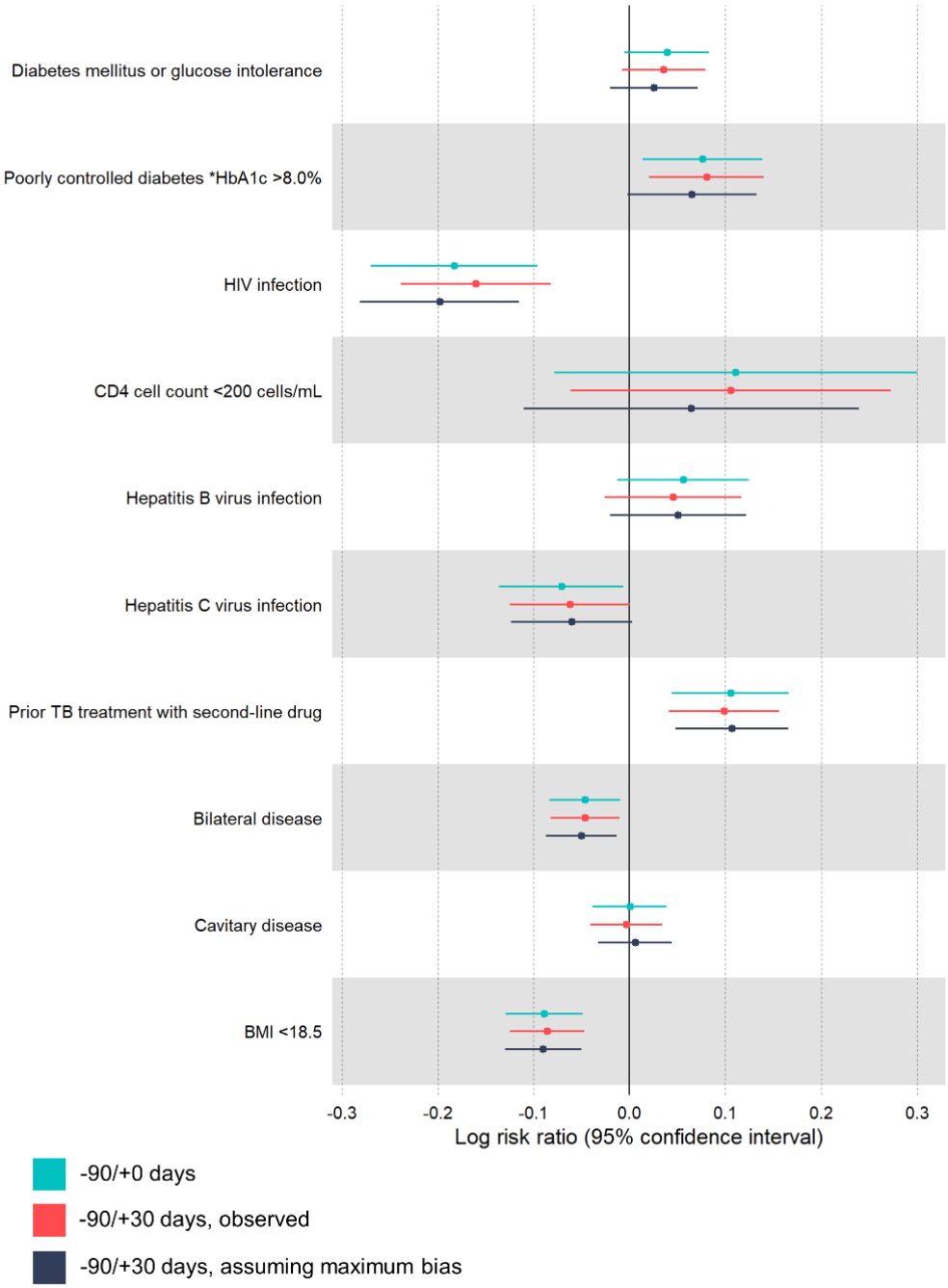


Figure 3.7 Crude relative risk of sputum culture conversion by baseline characteristics, comparison extending the interval 60 days past treatment initiation to no extension of interval past treatment initiation

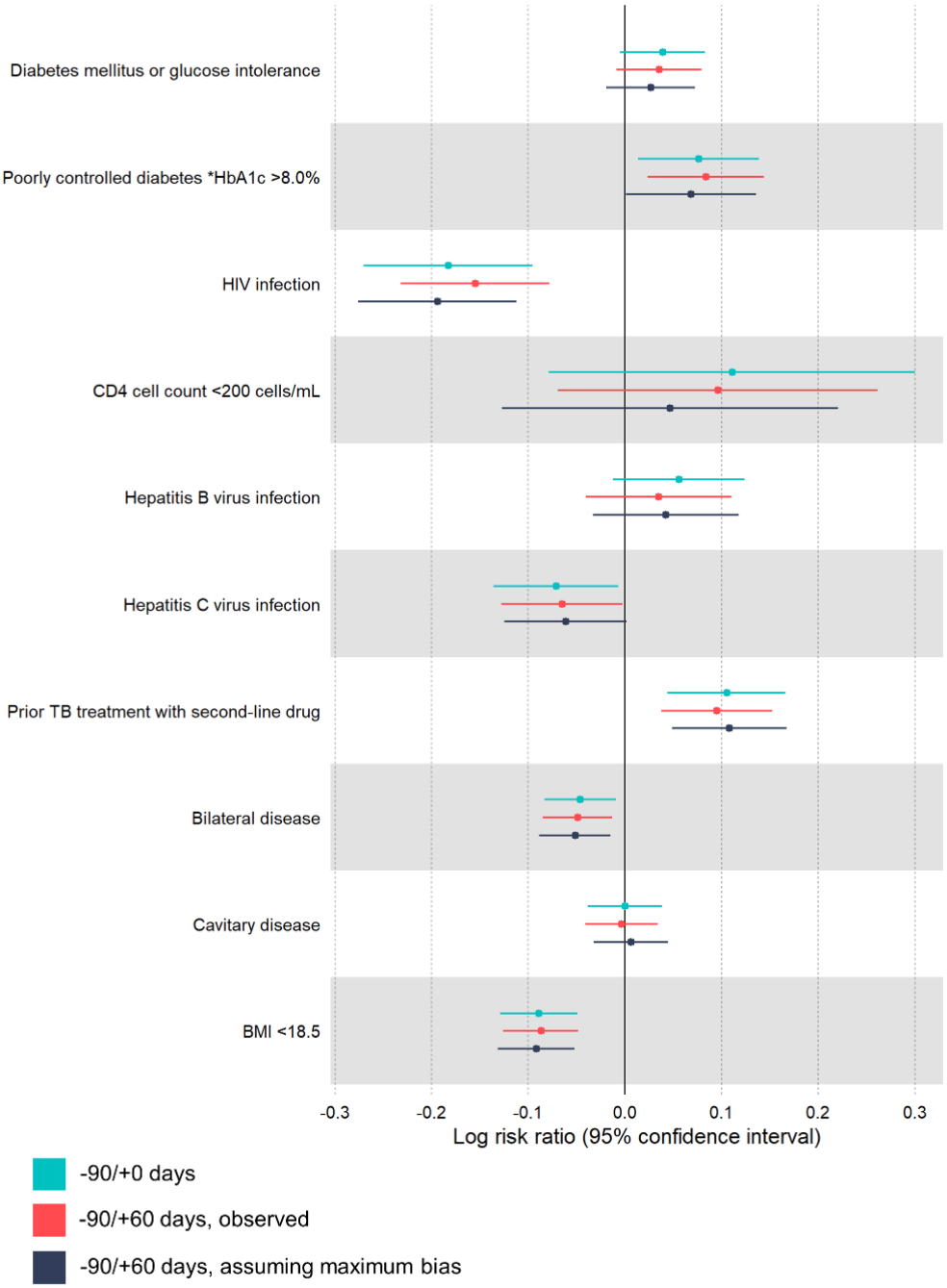


Figure 3.8 Crude relative risk of sputum culture conversion by baseline characteristics, comparison extending the interval 90 days past treatment initiation to no extension of interval past treatment initiation

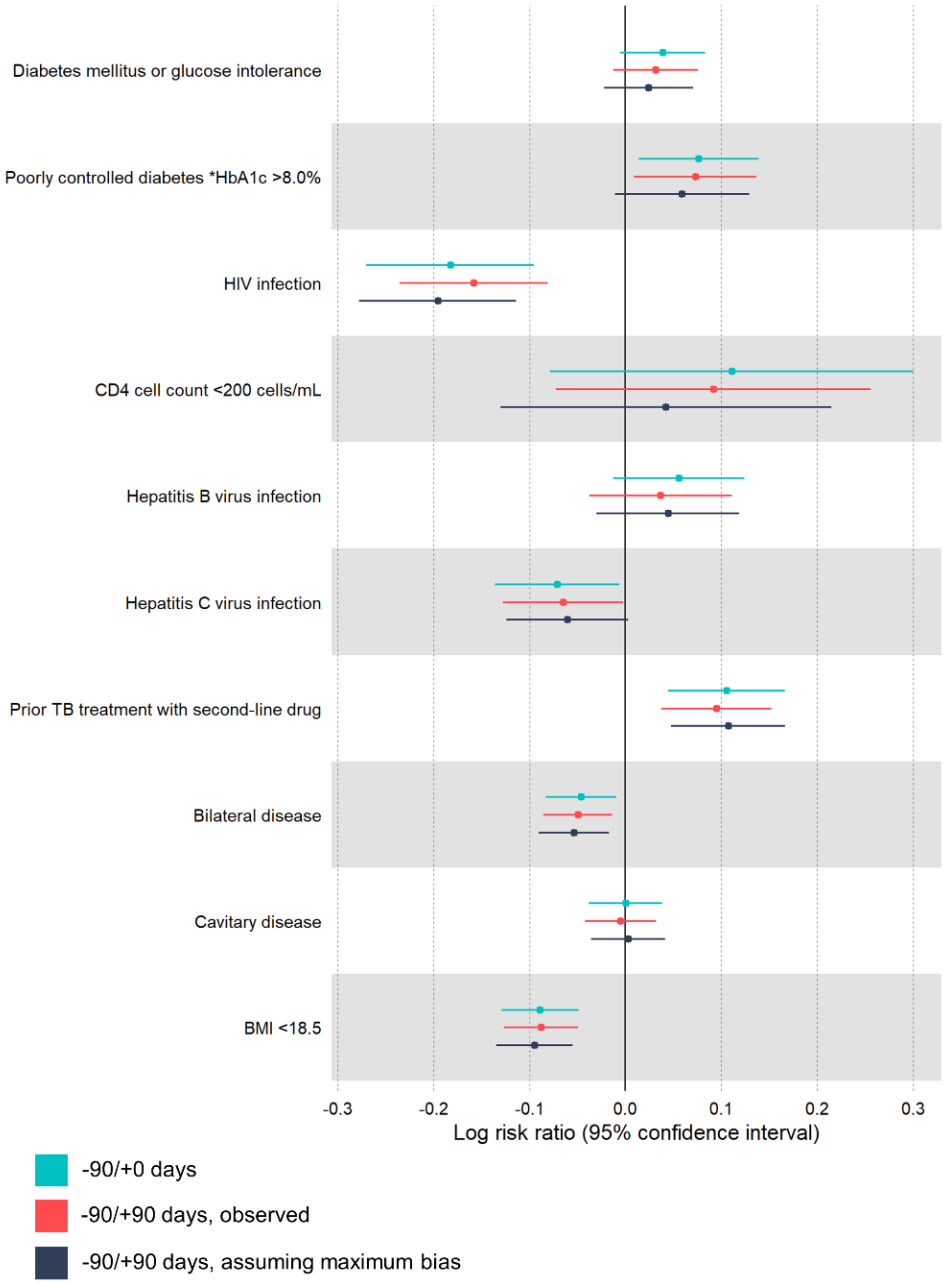


Table 3.4 Bias and precision in the relative risk of sputum culture conversion, comparison extending the interval 30 days past treatment initiation to no extension of interval past treatment initiation

Variable Baseline definition	N	Risk ratio (95% CI)	% bias[a]	MSE
Diabetes mellitus or glucose intolerance				
-90/+0 days	1687	1.04 (0.99, 1.09)	0	0.001
-90/+30 days, observed	1795	1.04 (0.99, 1.08)	-0.327	0.001
-90/+30 days, assuming maximum bias	1804	1.03 (0.98, 1.07)	-1.371	0.001
Poorly controlled diabetes (HbA1c >8.0%)				
-90/+0 days	861	1.08 (1.01, 1.15)	0	0.001
-90/+30 days, observed	905	1.08 (1.02, 1.15)	0.391	0.001
-90/+30 days, assuming maximum bias	912	1.07 (1, 1.14)	-1.117	0.001
HIV infection				
-90/+0 days	1690	0.83 (0.76, 0.91)	0	0.002
-90/+30 days, observed	1794	0.85 (0.79, 0.92)	2.257	0.002
-90/+30 days, assuming maximum bias	1809	0.82 (0.76, 0.89)	-1.517	0.002
CD4 cell count <200 cells/mL				
-90/+0 days	155	1.12 (0.92, 1.35)	0	0.009
-90/+30 days, observed	181	1.11 (0.94, 1.31)	-0.512	0.007
-90/+30 days, assuming maximum bias	187	1.07 (0.9, 1.27)	-4.515	0.01
Hepatitis B virus infection				
-90/+0 days	1703	1.06 (0.99, 1.13)	0	0.001
-90/+30 days, observed	1812	1.05 (0.97, 1.12)	-1.028	0.001

-90/+30 days, assuming maximum bias	1821	1.05 (0.98, 1.13)	-0.514	0.001
Hepatitis C virus infection				
-90/+0 days	1730	0.93 (0.87, 0.99)	0	0.001
-90/+30 days, observed	1841	0.94 (0.88, 1)	0.924	0.001
-90/+30 days, assuming maximum bias	1852	0.94 (0.88, 1)	1.107	0.001
Prior TB treatment with second-line drugs				
-90/+0 days	1496	1.11 (1.05, 1.18)	0	0.001
-90/+30 days, observed	1590	1.1 (1.04, 1.17)	-0.664	0.001
-90/+30 days, assuming maximum bias	1599	1.11 (1.05, 1.18)	0.119	0.001
Bilateral disease				
-90/+0 days	1519	0.95 (0.92, 0.99)	0	0
-90/+30 days, observed	1604	0.95 (0.92, 0.99)	-0.012	0
-90/+30 days, assuming maximum bias	1615	0.95 (0.92, 0.99)	-0.393	0
Cavitary disease				
-90/+0 days	1489	1 (0.96, 1.04)	0	0
-90/+30 days, observed	1571	1 (0.96, 1.04)	-0.318	0
-90/+30 days, assuming maximum bias	1579	1.01 (0.97, 1.05)	0.574	0
BMI <18.5				
-90/+0 days	1745	0.91 (0.88, 0.95)	0	0
-90/+30 days, observed	1853	0.92 (0.88, 0.95)	0.335	0
-90/+30 days, assuming maximum bias	1867	0.91 (0.88, 0.95)	-0.089	0

[a] $\frac{\exp(\beta_{-90/+\geq 1}) - \exp(\beta_{-90/+0})}{\exp(\beta_{-90/+0})} \times 100$

Table 3.5 Bias and precision in the relative risk of sputum culture conversion, comparison extending the interval 60 days past treatment initiation to no extension of interval past treatment initiation

Variable Baseline definition	N	Risk ratio (95% CI)	% bias[a]	MSE
Diabetes mellitus or glucose intolerance				
-90/+0 days	1687	1.04 (0.99, 1.09)	0	0.001
-90/+60 days, observed	1827	1.04 (0.99, 1.08)	-0.357	0.001
-90/+60 days, assuming maximum bias	1839	1.03 (0.98, 1.08)	-1.202	0.001
Poorly controlled diabetes (HbA1c >8.0%)				
-90/+0 days	861	1.08 (1.01, 1.15)	0	0.001
-90/+60 days, observed	916	1.09 (1.02, 1.15)	0.745	0.001
-90/+60 days, assuming maximum bias	923	1.07 (1, 1.15)	-0.777	0.001
HIV infection				
-90/+0 days	1690	0.83 (0.76, 0.91)	0	0.002
-90/+60 days, observed	1826	0.86 (0.79, 0.93)	2.837	0.002
-90/+60 days, assuming maximum bias	1845	0.82 (0.76, 0.89)	-1.103	0.002
CD4 cell count <200 cells/mL				
-90/+0 days	155	1.12 (0.92, 1.35)	0	0.009
-90/+60 days, observed	185	1.1 (0.93, 1.3)	-1.455	0.007
-90/+60 days, assuming maximum bias	192	1.05 (0.88, 1.25)	-6.21	0.012
Hepatitis B virus infection				
-90/+0 days	1703	1.06 (0.99, 1.13)	0	0.001
-90/+60 days, observed	1844	1.04 (0.96, 1.12)	-2.074	0.002
-90/+60 days, assuming maximum bias	1857	1.04 (0.97, 1.12)	-1.353	0.002

Hepatitis C virus infection				
-90/+0 days	1730	0.93 (0.87, 0.99)	0	0.001
-90/+60 days, observed	1873	0.94 (0.88, 1)	0.625	0.001
-90/+60 days, assuming maximum bias	1887	0.94 (0.88, 1)	0.992	0.001
Prior TB treatment with second-line drugs				
-90/+0 days	1496	1.11 (1.05, 1.18)	0	0.001
-90/+60 days, observed	1615	1.1 (1.04, 1.16)	-1.015	0.001
-90/+60 days, assuming maximum bias	1627	1.11 (1.05, 1.18)	0.277	0.001
Bilateral disease				
-90/+0 days	1519	0.95 (0.92, 0.99)	0	0
-90/+60 days, observed	1630	0.95 (0.92, 0.99)	-0.272	0
-90/+60 days, assuming maximum bias	1643	0.95 (0.92, 0.99)	-0.526	0
Cavitary disease				
-90/+0 days	1489	1 (0.96, 1.04)	0	0
-90/+60 days, observed	1595	1 (0.96, 1.03)	-0.351	0
-90/+60 days, assuming maximum bias	1605	1.01 (0.97, 1.05)	0.615	0
BMI <18.5				
-90/+0 days	1745	0.91 (0.88, 0.95)	0	0
-90/+60 days, observed	1883	0.92 (0.88, 0.95)	0.21	0
-90/+60 days, assuming maximum bias	1901	0.91 (0.88, 0.95)	-0.263	0

[a] $\frac{\exp(\beta_{-90/+\geq 1}) - \exp(\beta_{-90/+0})}{\exp(\beta_{-90/+0})} \times 100$

Table 3.6 Bias and precision in the relative risk of sputum culture conversion, comparison extending the interval 90 days past treatment initiation to no extension of interval past treatment initiation

Variable Baseline definition	N	Risk ratio (95% CI)	% bias[a]	MSE
Diabetes mellitus or glucose intolerance				
-90/+0 days	1687	1.04 (0.99, 1.09)	0	0.001
-90/+90 days, observed	1834	1.03 (0.99, 1.08)	-0.76	0.001
-90/+90 days, assuming maximum bias	1848	1.02 (0.98, 1.07)	-1.471	0.001
Poorly controlled diabetes (HbA1c >8.0%)				
-90/+0 days	861	1.08 (1.01, 1.15)	0	0.001
-90/+90 days, observed	918	1.08 (1.01, 1.15)	-0.346	0.001
-90/+90 days, assuming maximum bias	926	1.06 (0.99, 1.14)	-1.712	0.002
HIV infection				
-90/+0 days	1690	0.83 (0.76, 0.91)	0	0.002
-90/+90 days, observed	1833	0.85 (0.79, 0.92)	2.479	0.002
-90/+90 days, assuming maximum bias	1854	0.82 (0.76, 0.89)	-1.288	0.002
CD4 cell count <200 cells/mL				
-90/+0 days	155	1.12 (0.92, 1.35)	0	0.009
-90/+90 days, observed	186	1.1 (0.93, 1.29)	-1.881	0.007
-90/+90 days, assuming maximum bias	193	1.04 (0.88, 1.24)	-6.627	0.012
Hepatitis B virus infection				
-90/+0 days	1703	1.06 (0.99, 1.13)	0	0.001
-90/+90 days, observed	1851	1.04 (0.96, 1.12)	-1.926	0.002
-90/+90 days, assuming maximum bias	1865	1.05 (0.97, 1.13)	-1.151	0.002

Hepatitis C virus infection				
-90/+0 days	1730	0.93 (0.87, 0.99)	0	0.001
-90/+90 days, observed	1880	0.94 (0.88, 1)	0.634	0.001
-90/+90 days, assuming maximum bias	1895	0.94 (0.88, 1)	1.059	0.001
Prior TB treatment with second-line drugs				
-90/+0 days	1496	1.11 (1.05, 1.18)	0	0.001
-90/+90 days, observed	1621	1.1 (1.04, 1.16)	-1.038	0.001
-90/+90 days, assuming maximum bias	1634	1.11 (1.05, 1.18)	0.177	0.001
Bilateral disease				
-90/+0 days	1519	0.95 (0.92, 0.99)	0	0
-90/+90 days, observed	1637	0.95 (0.92, 0.99)	-0.346	0
-90/+90 days, assuming maximum bias	1652	0.95 (0.91, 0.98)	-0.769	0
Cavitary disease				
-90/+0 days	1489	1 (0.96, 1.04)	0	0
-90/+90 days, observed	1602	1 (0.96, 1.03)	-0.514	0
-90/+90 days, assuming maximum bias	1614	1 (0.96, 1.04)	0.251	0
BMI <18.5				
-90/+0 days	1745	0.91 (0.88, 0.95)	0	0
-90/+90 days, observed	1890	0.92 (0.88, 0.95)	0.136	0
-90/+90 days, assuming maximum bias	1910	0.91 (0.87, 0.95)	-0.573	0

[a] $\frac{\exp(\beta_{-90/+\geq 1}) - \exp(\beta_{-90/+0})}{\exp(\beta_{-90/+0})} \times 100$

4 COMPARATIVE EFFECTIVENESS STUDIES OF TIME-VARYING TREATMENTS IN MULTIDRUG-RESISTANT TUBERCULOSIS TREATMENT COHORTS

4.1 Introduction

Conventional, longer regimens used to treat MDR-TB are administered for up to 24 months and are often individualized to patient characteristics at treatment initiation. These baseline patient characteristics used to determine the initial treatment regimen may also be associated with the outcome, a common source of bias in observational cohort studies known as confounding.(85) Regimens can also change over the course of treatment based on patients' evolving clinical status. Consequently, MDR-TB treatment is, for many patients, a time-varying exposure.(9) Time-varying treatments can be subject to treatment-confounder feedback, known as time-dependent confounding affected by previous exposure.(41,86) Time-dependent confounding occurs when a time-varying factor is associated with the outcome, predicts subsequent treatment, and is affected by past treatment. For example, the regimen administered at treatment initiation, referred to as "baseline", is often individualized based on comorbidities, such as HIV, and extent of drug resistance. A provider may subsequently change the baseline regimen based on time-varying factors, such as development of further drug resistance or lack of microbiological improvement (i.e. conversion of sputum smear and culture). These time-varying factors strongly predict treatment outcome and are affected themselves by

previous treatment.(42) Methods to control for time-dependent confounding include inverse probability weighting in marginal structural models, the g-formula, and structural nested models.(19,41) Based on a 2020 systematic search for longitudinal studies of TB treatment cohorts using these methods to adjust for time-dependent confounding, we identified that none of these methods have been applied to TB-focused research questions.(42)

Since most observational analyses of MDR-TB treatment classify patients' regimens only according to their composition at treatment initiation (baseline), regardless of whether treatment changed over the course of follow up, they fail to account for the fact that MDR-TB treatment is time-varying. Comparative effectiveness analyses using baseline treatment definitions produce estimates of the observational analogue of the intention-to-treat effect, that is, the effect of *initiating* a particular drug or regimen. Often, of greater policy and clinical interest is the observational analogue of the per-protocol effect, which estimates the effect of *initiating and completing a treatment course* with a particular drug or regimen for its intended duration.(81,87,88) In fact, many of the priority research questions identified by WHO in the 2019 MDR-TB treatment guidelines focus on regimen composition and length of treatment,(61) highlighting the importance of estimating per-protocol effects.

Analyses of the per-protocol analogue have the added challenge of adjustment for time-dependent confounding. One method to estimate per-protocol effects and control for time-dependent confounding is to artificially censor patients' follow up time when their treatment deviates from the treatment administered at baseline and use inverse probability

of censoring weights to adjust for selection bias due to artificial censoring. Through weighting, this approach creates a pseudo-population where all individuals complete their initial treatment and where time-varying treatment and time-varying confounders are no longer associated.(43) However, in order to implement this analytic approach, time-varying data on predictors of treatment regimen changes must be measured. To date, few cohorts have collected the extent of longitudinal data need to adjust for bias, precluding the ability to estimate per-protocol effects in the observational context. When data are available to conduct both the intention-to-treat and per-protocol analogue, investigators should pursue analyses that produce estimates of both and compare findings. Differences between the analyses are dependent on the extent of protocol deviations (e.g. treatment regimen changes in MDR-TB cohorts). When deviations are rare, differences between analyses are likely to be minimal. Conversely, differences may be larger when deviations are common and can provide insight into the presence of time-dependent confounding.

In Aim 1, we applied the aforementioned artificial censoring approach to answer the research question of whether adding delamanid to an MDR-TB regimen comprised of only three drugs likely to be effective provided a benefit on two- and six-month sputum culture conversion. Here, for this same research question, we (1) compare weighted and unweighted models estimating the per-protocol analogue to assess the impact of selection bias (and, by extension, of time-dependent confounding) and (2) assess whether estimates of the observational analogue of the per-protocol and intention-to-treat meaningfully differ.

4.2 **Methods**

4.2.1 Data source, study population, and definitions

We used the same data source and study population as in Aim 1. In brief, we included a subset of participants from the endTB observational cohort who had a positive baseline sputum culture and received one of the following baseline MDR-TB regimens of interest: (1) "delamanid-containing" regimens comprised of delamanid plus a background regimen of three drugs likely to be effective, (2) "delamanid-free" regimens comprised of three drugs likely to be effective, none of which was delamanid. We assessed the relative risk and risk difference of two- and six-month culture conversion across groups using the same outcomes definitions described in Aim 1.

4.2.2 Marginal versus conditional effects

In order to make comparisons across models producing estimates with analogous interpretations, we divide results by whether the selected model produces estimates of marginal or conditional effects. Marginal models produce estimates of the average effect at the *population level* of moving the population from unexposed to exposed. Conditional models produce estimates of average effect that are relevant within strata defined by a set of baseline covariates.(89)

4.2.2.1 Descriptive analysis of the distribution of censoring

To investigate the degree to which exposure varied over time, we assessed the distribution of censoring in the data according to baseline exposure group and each outcome. We additionally calculated crude relative risks to identify the association

between censoring and the exposure (delamanid-containing regimens) and each outcome using the following formula, where C=1 represents the number of participants censored and Y=1 represents the number of participants on delamanid-containing regimens or that converted: $\frac{C=1,Y=1}{(C=1,Y=1)+(C=0,Y=1)} / \frac{C=1,Y=0}{(C=1,Y=0)+(C=0,Y=0)}$.

4.2.3 Inverse probability censoring weighted analysis of the per-protocol analogue

We used the censoring approach described at length in Aim 1 to estimate the per-protocol analogue. This analysis produces an estimate of the per-protocol analogue representing the effect of initiating an intervention and maintaining the intervention for a protocol-defined duration. In brief, we artificially censored participants if the addition or subtraction of a drug resulted in an exposure group switch lasting >2 weeks. Artificial censoring will introduce selection bias, which must be controlled for via inverse probability of censoring weights. To calculate inverse probability of censoring weights, we fitted a pooled logistic regression to estimate the probability that the participant was not censored (i.e. that their initial exposure group would be sustained, conditional on time-varying predictors of changing treatment). We calculated weights by multiplying the probability of being uncensored across each time point. Control for baseline confounders of treatment differed, dependent on whether the intention was to estimate marginal (Section 4.2.3.1) or conditional effects (Section 4.2.3.2).

Only participants whose initial treatment was unchanged (i.e. uncensored participants) were included in the final logistic model with their associated weights. Using the same sub-cohort of uncensored participants, we compared estimates from models that control

(weighted analysis) versus those that do not control (unweighted analysis) for selection bias and, by extension, time-dependent confounding. We qualitatively describe the difference between these estimates.

4.2.3.1 Marginal effects of the censoring weighted per-protocol analogue

We used composite inverse probability of treatment weights and inverse probability of censoring weights to estimate the marginal average treatment effect of delamanid-containing regimens versus delamanid-free regimens.

To calculate inverse probability of treatment weights, all participants were assigned a time-fixed weight defined as the inverse of the probability of receiving the treatment actually received conditional on baseline confounders of treatment. Through content knowledge and directed acyclic graphs, we selected confounders and fitted multiple baseline adjusted models, as described in Aim 1, Appendix 5.4. We selected a final baseline model comprised of age, sex, whether the participant was in the hospital at treatment initiation, the number of Group A drugs in the regimen, whether the patient was on imipenem-cilastatin, body mass index <18.5, HIV infection, and hepatitis C infection. The following formula represents the unstabilized treatment weight (W^A) for exposed participants, where A is the baseline exposure group (1=delamanid-containing, 0=delamanid-free) and L is the vector of baseline covariates: $W^A = \frac{1}{\Pr[A=1 \mid L]}$. For unexposed participants, W^A is represented as $W^A = \frac{1}{1-\Pr[A=1 \mid L]}$. Inverse probability of treatment weighting creates a pseudopopulation where treatment and confounders are unrelated to each other.

Inverse probability of censoring weights, described in Section 4.2.3 and treatment weights were multiplied together to calculate a composite weight, as represented by the formula $W^C(t) = \prod_{k=0}^{t} \frac{1}{\Pr\left[C(k) = 0 \middle| L, \bar{L}(k),\ C(k-1) = 0,\ A\right] \times \Pr[A = 1|L]}$ for exposed participants and $\prod_{k=0}^{t} \frac{1}{\Pr\left[C(k) = 0 \middle| L, \bar{L}(k),\ C(k-1) = 0,\ A\right] \times 1 - \Pr[A = 1|L]}$ for unexposed participants. We fitted a second logistic regression model including uncensored participants weighted by their composite weight. From this model, we calculated the predicted probabilities of two- and six-month culture conversion for each participant. Then, we computed the mean predicted probability by exposure group, and used the probabilities of conversion by exposure group to calculate the relative risk and risk difference.(90) Confidence intervals were estimated using nonparametric bootstrapping with 500 samples.

4.2.3.2 Conditional effects of the censoring weighted per-protocol analogue

To estimate conditional effects of the per-protocol analogue, we applied inverse probability of censoring weights to uncensored participants in a second logistic regression model and directly adjusted for baseline confounders in the model. This direct baseline adjustment method is in contrast to the inclusion of inverse probability of treatment weights used in the marginal model. Relative risks and risk differences were calculated using the mean predicted probability and nonparametric bootstrapping.(90)

4.2.4 Baseline-adjusted analysis of the intention-to-treat analogue

The observational analogue of the intention-to-treat effect estimates the effect of

initiating an intervention and does not account for post-baseline changes in treatment. Thus, we classified participants according to the treatment they had received at baseline (defined as day 7 of treatment) and adjusted for baseline confounders of treatment using several methods, described here in Section 4.2.4.1 and Section 4.2.4.2:

4.2.4.1 Marginal effects of the baseline-adjusted intention-to-treat analogue

For marginal effects of the intention-to-treat analogue, we estimated baseline treatment weights using the same approach described in Section 4.2.3.1 above. We fitted a weighted logistic regression model with culture conversion as the dependent variable, treatment group as the independent variable and treatment weights W^A.

4.2.4.2 Conditional effects of the baseline-adjusted intention-to-treat analogue

We used logistic regression and log binomial regression to estimate conditional effects of the intention-to-treat analogue. Both regression methods produce estimates of the association of treatment and the outcome within levels of the baseline confounders in the model.

We fitted a logistic regression model on baseline confounders and estimated risks of conversion by exposure group using the predicted probabilities from the model and nonparametric bootstrapping to estimate confidence intervals.(90) The log binomial model directly estimates relative risks,(91) thus indirect estimation of risks was not necessary. Confidence intervals from the log binomial model were estimated using robust standard errors.

4.3 **Results**

4.3.1 Inverse probability censoring weighted analysis of the per-protocol analogue

Analyses of the per-protocol analogue include 349 participants: 10 participants were excluded for missing time-varying data, two for missing both time-varying data and baseline data, and two for missing baseline data (Table 4.1).

4.3.1.1 Distribution of censoring

As reported in Aim 1, 97 (27%) participants were artificially censored because their regimen was changed such that it resulted in an exposure group switch for more than two weeks. Censoring occurred at a median of 11 weeks (25^{th} percentile=5 weeks, 75^{th} percentile=17 weeks). Censored participants were proportionally represented within their respective baseline exposure groups (28.8% censored in delamanid-containing, 25.6% censored in delamanid-free) (Table 4.2). Receiving a delamanid-containing regimen was not strongly associated with changing exposure groups (i.e. censoring) (crude RR: 1.12 (95% CI: 0.79, 1.60) (Table 4.3). Participants who converted were less likely to have changed exposure groups (i.e. be censored) than those who did not experience conversion (crude RR at two months: 0.63 (95% CI: 0.45, 0.89); crude RR at six months: 0.66 (95% CI: 0.44, 0.98)) (Table 4.3).

4.3.1.2 Marginal effects of the censoring weighted per-protocol analogue

Using composite inverse probability of treatment and censoring weights resulted in a relative risk of conversion at two months of 0.93 (95% CI: 0.52, 2.36) and 0.90 (95% CI: 0.56, 1.76) at six months (Table 4.1, Model 3).

4.3.1.3 Conditional effects of the censoring weighted per-protocol analogue

Using inverse probability of censoring weights and adjusting for baseline confounders in the final model, participants receiving delamanid had a lower risk of two-month conversion, albeit with a confidence interval that included the null (RR: 0.93 (95% CI: 0.67, 1.41)) (Table 4.1, Model 5). Similar results were observed for six month conversion.

4.3.1.4 Comparison of weighted and unweighted models to assess impact of selection bias

Across models estimating marginal and conditional effects, weighted and unweighted point estimates were similar (Table 4.1, Model 2 vs 3 and Model 4 vs 5). Confidence intervals from weighted analyses were notably wider than those that omitted the weights. Similar results were observed for risk differences.

4.3.2 Baseline-adjusted intention-to-treat analogue

Among the subcohort of 363 participants identified in Aim 1, 359 participants had complete data on baseline confounders and were included in baseline-adjusted analyses of the intention-to-treat analogue (Table 4.1).

4.3.2.1 Marginal effects of the baseline-adjusted intention-to-treat analogue

Using inverse probability of treatment weights, participants on delamanid-containing regimens had a lower risk of culture conversion at two months when compared to participants on delamanid-free regimens (RR: 0.85 (95% CI: 0.66, 1.08), RD: -0.08 (95% CI: -0.19, 0.04)) (Table 4.1, Model 6). Estimates for six month culture conversion were

slightly attenuated towards the null (RR: 0.88 (95% CI: 0.80, 0.98), RD: -0.10 (95% CI: -0.17, 0.02)) (Table 4.1, Model 6).

4.3.2.2 *Conditional effects of the baseline-adjusted intention-to-treat analogue*

Using the baseline-adjusted logistic regression approach, the relative risk of culture conversion was 0.90 (95% CI: 0.71, 1.13) at two months and 0.93 (95% CI: 0.84, 1.01) at six months (Table 4.1, Model 7). The log binomial model yielded similar relative risks, with point estimates slightly closer to the null (Table 4.1, Model 8).

4.3.3 Observational analogue of the intention-to-treat versus per protocol analyses

Despite fluctuations in point estimates, estimates of the intention-to-treat and per protocol analogue consistently yielded similar results. A minor exception to this is that the marginal effect of the intention-to-treat analogue (Table 4.1, Model 6) indicated participants on delamanid-containing regimens had relatively less conversion than participants on delamanid-free regimens by two months (52% vs 56%, RR: 0.85 (95% CI: 0.66, 1.08) in comparison to the per-protocol analogue (46% vs 54%, RR: 0.93 (95% CI: 0.52, 2.36) (Table 4.1, Model 3). However, the variability around the per-protocol analogue estimate was larger. Conditional effects of the intention-to-treat analogue (Table 4.1, Models 7 and 8) were nearly identical to that of the per protocol analogue (Table 4.1, Model 5).

4.4 **Discussion**

Using models accounting for the time-varying nature of MDR-TB treatment did not change our conclusion that there was no effect of adding delamanid to MDR-TB regimens containing three drugs likely to be effective, despite modest fluctuations in point estimates and confidence intervals across models. Estimates of the observational analogue of the per-protocol effect and the intention-to-treat effect were similar. Both of these findings are likely due to the frequency and distribution of censoring (i.e. exposure group changes) in our data for the research question at hand.

We did not identify a difference between censoring weighted and unweighted per-protocol analyses. Participants were censored if their exposure group changed, either due to the addition or subtraction of delamanid or a drug change that resulted in the participant no longer receiving three drugs likely to be effective in the background regimen. The likely reason for not identifying differences between these analyses is that changes to the primary drug of interest, delamanid, were rare (12/97, 12%). Artificial censoring was more common for changes in the background regimen, such that they were no longer receiving three background drugs likely to be effective. Importantly, a necessary condition for there to be meaningful differences between estimates is that censoring be highly associated with the exposure and the outcome.(82,92) We cannot rule out that misspecification of the model used to derive censoring weights could explain our finding that there were not meaningful differences between weighted and unweighted models. However, we tested multiple models with different variable combinations and specifications and observed similar results (Appendices 5.2 and 5.3).

We also did not identify a clinically meaningful difference between results of standard baseline-adjusted models (logistic regression, log binomial regression), which estimate the intention-to-treat analogue, and the inverse probability censoring weighted analysis estimating the per-protocol analogue. For this particular research question, the effect of adding delamanid to a three drug regimen at baseline (i.e. intention-to-treat analogue) is similar to that of adding and maintaining delamanid in a three drug regimen for the first six months of treatment (i.e. per-protocol analogue). Like that of the weighted and unweighted per-protocol analyses, this may be due to the small proportion of participants censored and censoring's distribution across exposure groups and the outcomes.

Despite our primary finding that results were relatively consistent across various models, investigators conducting research with MDR-TB treatment cohorts should still pursue analyses that consider the time-varying nature of MDR-TB treatment. This approach is particularly important when treatment changes are driven by factors that also impact treatment outcome. For example, a research priority highlighted in the 2019 MDR-TB guidelines is the optimization of the dose and duration of linezolid.(61) Linezolid is a highly active anti-TB drug that often results in significant toxicity.(9,93–97) When toxicity occurs, such as the development of myelosuppression or peripheral neuropathy, cessation of linezolid is recommended. A comparative safety observational study assessing the effect of three months of linezolid versus six months of linezolid on a safety endpoint will be biased by the fact that patients receive shorter durations of linezolid *because of* toxicity. In such an analysis, continuation of linezolid for six months would be highly associated with the absence of toxicity (the outcome) in previous months. More

complex analyses using censoring and weighting techniques are required to control for bias.(98)

The present study highlights some important analytic issues that investigators working with MDR-TB treatment cohorts should consider. Baseline adjusted comparative effectiveness analyses are the standard in assessing efficacy of MDR-TB treatment regimens.(23,58) Depending on the research question, these analyses will not necessarily be biased. However, they may not be answering the appropriate clinical question if the intention-to-treat analogue is not of greatest interest for decision makers. Rather, if the objective is to estimate the per-protocol analogue, the artificial censoring technique applied here is a simple way to identify the effect of treatment for a protocolized duration while simultaneously resolving the potential for time-dependent confounding.(43) Investigators should carefully define the most important research questions, identify the exposure groups to be compared, and use an analytic approach that will produce an estimate of the causal effect (intention-to-treat or per-protocol) that is intended. When possible, investigators can conduct analyses to assess both the intention-to-treat and per-protocol analogues. Target trial emulation is an intuitive framework to assist investigators through these steps.(35,37)

A critical barrier to conducting analyses that account for treatment changes is the absence of data of sufficient quality and granularity. However, a major strength of this study is that the endTB cohort is the first longitudinal study of its size to have systematically collected prospective data on factors that have otherwise been absent in other programmatic cohorts. Through monthly patient encounters and intensive

pharmacovigilance, detailed data on these potential confounders have been collected and can be controlled for appropriately. Without such data, analyses accounting for the time-varying nature of treatment would not be possible. The limitations of this study are those that are inherent in using observational data to estimate causal effects. As previously noted, we cannot rule out the potential for model misspecification or unmeasured confounding; if these occurred, the observed associational estimates would not equate to causal effects. However, given that the study could draw from extensive longitudinal data and that it produced consistent findings across multiple models, the risk of residual bias due to model misspecification and unmeasured confounding is likely low.

Using the clinical question of adding delamanid to a three drug MDR-TB regimen as a guide, we determined, across multiple models with different interpretations, that the addition of delamanid did not have a significant effect on culture conversion at either of two time points. Treatment regimen changes were largely unrelated to delamanid and did not predict culture conversion, which likely explains the consistency of estimates. Despite these findings, investigators should still implement analyses that consider the time-varying nature of MDR-TB treatment, especially when the clinical question of interest involves a treatment that frequently changes and when treatment changes are highly associated with one exposure group and the outcome.

4.5 Tables and Figures

Table 4.1 Comparison of analyses that assess the effect of adding delamanid to a regimen composed of three drugs likely to be effective, endTB observational cohort (N=363)

Model #	Analysis	N	Two-month culture conversion		Six-month culture conversion	
			RR (95% CI)	RD (95% CI)	RR (95% CI)	RD (95% CI)
1	Crude	363	0.89 (0.72, 1.09)	-0.06 (-0.17, 0.05)	0.91 (0.82, 1.00)	-0.08 (-0.16, -0.00)
IP censoring weighted (unstabilized), per protocol analogue						
	Marginal effects					
2	Unweighted (biased)	349[a]	0.90 (0.71, 1.16)	-0.06 (-0.19, 0.09)	0.93 (0.82, 1.02)	-0.07 (-0.17, 0.01)
3	Weighted[b]	349[a]	0.93 (0.52, 2.36)	-0.04 (-0.30, 0.40)	0.90 (0.56, 1.76)	-0.09 (-0.42, 0.38)
	Conditional effects					
4	Unweighted (biased)	349[a]	0.90 (0.71, 1.13)	-0.06 (-0.18, 0.08)	0.92 (0.82, 1.01)	-0.07 (-0.16, 0.01)
5	Weighted	349[a]	0.93 (0.67, 1.41)	-0.04 (-0.20, 0.20)	0.93 (0.78, 1.05)	-0.06 (-0.21, 0.05)
Baseline adjusted, intention-to-treat analogue						
	Marginal effects					
6	IP treatment weighted (unstabilized)[d]	359[c]	0.85 (0.66, 1.08)	-0.08 (-0.19, 0.04)	0.88 (0.80, 0.98)	-0.10 (-0.17, -0.02)

	Conditional effects					
7	Logistic regression	359[c]	0.90 (0.71, 1.13)	-0.05 (-0.17, 0.07)	0.93 (0.84, 1.01)	-0.06 (-0.15, 0.01)
8	Log binomial[e]	359[c]	0.92 (0.74, 1.13)	-	0.97 (0.87, 1.08)	-

Abbreviations: Inverse probability (IP), risk ratio (RR), risk difference (RD), bedaquiline (BDQ)

[a] N=10 participants excluded for missing time-varying data, N=2 participants excluded for missing time-varying and baseline data, N=2 participants excluded for missing baseline data

[b] Unstabilized IP weight (product of censoring and treatment) mean (minimum, maximum): 2.56 (1.06, 18.6)

[c] N=4 participants excluded for missing baseline data

[d] Unstabilized IP treatment weight mean (minimum, maximum): 1.96 (1.01, 13.71)

[e] Log binomial regression does not produce estimates of risk differences

Table 4.2 Proportion of participants censored, by exposure group and two- and six-month culture conversion in analysis to assess the effect of adding delamanid to a regimen composed of three drugs likely to be effective, endTB observational cohort (N=363)

	Delamanid-containing (N=125)		Delamanid-free (N=238)	
	Censored	Uncensored	Censored	Uncensored
Total, n/N (%)	36/125 (28.8)	89/125 (71.2)	61/238 (25.6)	177/238 (74.4)
Two-month, converted, n/N (%)	14/36 (38.9)	48/89 (53.9)	27/61 (44.3)	106/177 (59.9)
Two-month, did not convert, n/N (%)	22/36 (61.1)	41/89 (46.1)	34/61 (55.7)	71/177 (40.1)
Six-month, converted, n/N (%)	28/36 (77.8)	73/89 (82.0)	50/61 (82.0)	162/177 (91.5)
Six-month, did not convert, n/N (%)	8/36 (22.2)	16/89 (18.0)	11/61 (18.0)	15/177 (8.5)

Table 4.3 Crude association of censoring with exposure group and outcome in analysis to assess the effect of adding delamanid to a regimen composed of three drugs likely to be effective, endTB observational cohort (N=363)

Association	Risk ratio (95% CI)
Association between censoring and exposure group	1.12 (0.79, 1.60)
Association between censoring and two-month conversion	0.63 (0.45, 0.89)
Association between censoring and six-month conversion	0.66 (0.44, 0.98)

5 APPENDIX

Appendix 5.1 Target trial protocol and emulation of the target trial using endTB observational cohort data

Objective: Evaluate the effect of adding delamanid to an MDR-TB regimen that contains 3 likely effective drugs.

	Target trial	Observational analysis (endTB observational cohort)
Eligibility criteria	**Inclusion** - Age 14+ - Diagnosis of MDR-TB from: 1) a sputum specimen collected no more than 90 days prior to the date of trial screening with positive growth of *M.tb* in culture and documented phenotypic resistance to rifampin OR 2) a sputum specimen collected no more than 90 days prior to the date of trial screening with positive growth of *M.tb* in culture, and a genotypic test positive for *M.tb* and with mutations known to confer resistance to rifampin	Same
Treatment strategies	**Intervention** Delamanid plus a background regimen of 3 likely effective drugs administered for 24 weeks, where a likely effective drug is defined as a drug for which resistance testing indicated participants' *M.tb* strain was not resistant to the drug or a drug for which no resistance testing was conducted and the participant had not previously received the drug for ≥1 month **Comparator** A regimen of 3 likely effective drugs, none of which are delamanid, administered for 24 weeks	Same

Treatment assignment	Randomized	Randomization assumed within levels of covariates
Follow up	**Time zero** Treatment assignment **End of follow up** Outcome, 24 weeks after time zero **LTFU** Treatment interruption for ≥2 months	**Time zero** The earliest start date of bedaquiline or delamanid initiation with endTB **End of follow up** Same **LTFU** Same
Outcome	Six-month culture conversion, defined as two, consecutive negative cultures collected at least 15 days apart, the first occurring before 180 days of treatment and the second before 210 days; death and LTFU prior to 180 days are considered non-conversion events.	Same
Causal contrast	Intention-to-treat effect	Per-protocol effect
Statistical analysis	Crude ratio and difference of the proportion with two-month and six-month culture conversion across arms **Main effect measure** Two-month culture conversion risk ratio and risk difference Six-month culture conversion risk ratio and risk difference	Inverse probability censoring weights **Main effect measure** Same

Appendix 5.2 Model specification and weight diagnostics for inverse probability of censoring weights

Model	Specification	Weight	N	Mean (SD)	Min, max
1	Numerator includes time (stabilized). Denominator includes treatment group, time-varying linear term for # Group A drugs, time-varying smear result (+/-), time-varying linear term for # SAEs/time, time-varying linear term for # AEs/time, time-varying term for hospitalization (0/1), baseline age and sex, a 3-knot spline for time, time.	Unstabilized	349[a]	1.35 (0.56)	1.00, 7.25
		Stabilized	349[a]	0.99 (0.40)	0.75, 5.24
2[b]	Numerator includes time (stabilized). Denominator includes exposure group, time-varying linear term for # Group A drugs, time-varying smear result (+/-), time-varying linear term for # AEs, time-varying term for hospitalization (0/1), a 3-knot spline for time, time.	Unstabilized	349[a]	1.34 (0.44)	1.00, 6.40
		Stabilized	349[a]	0.98 (0.31)	0.76, 4.62
3	Numerator includes time (stabilized). Denominator includes exposure group, time-varying linear term for # Group A drugs, time-varying smear result (+/-), time-varying linear term for # AEs/time, a 3-knot spline for time, time.	Unstabilized	349[a]	1.34 (0.48)	1.00, 7.37
		Stabilized	349[a]	0.98 (0.34)	0.76, 5.33
4	Numerator includes time (stabilized). Denominator includes exposure group, time-varying linear term for # Group A drugs, time-varying linear term for number of SAEs/time, time-varying linear term for # AEs/time, time-varying term for hospitalization (0/1), baseline age and sex, a 3-knot spline for time, time.	Unstabilized	359[c]	1.35 (0.56)	1.00, 7.86
		Stabilized	359[c]	0.99 (0.40)	0.75, 5.68

BDQ	Numerator includes time (stabilized). Denominator includes exposure group, time-varying linear term for # Group A drugs, time-varying smear result (+/-), time-varying linear term for # AEs, time-varying term for hospitalization (0/1), a 3-knot spline for time, time	Unstabilized	288[d]	1.30 (0.35)	1.00, 4.04
		Stabilized	288[d]	0.98 (0.30)	0.77, 3.02

Abbreviations: Standard deviation (SD), adverse event (AE), serious adverse event (SAE), Sensitivity analysis restricted to participants also receiving bedaquiline (BDQ)

[a] N=10 participants excluded for missing time-varying data, N=2 participants excluded for missing baseline and time-varying data, N=2 participants excluded for missing baseline data

[b] Model used in primary analysis

[c] N=7 participants excluded for missing time-varying data N=1 participant excluded for missing baseline data

[d] N=4 participants excluded for missing baseline data

Appendix 5.3 Mean predicted probabilities, mean risk ratio, and mean risk difference, by model specification

Model	Weight	N	Two-month culture conversion		Six-month culture conversion	
			Mean RR	Mean RD	Mean RR	Mean RD
1	Unstabilized	349[a]	0.97	-0.02	0.94	-0.06
	Stabilized	349[a]	0.97	-0.02	0.94	-0.06
2[b]	Unstabilized	349[a]	0.94	-0.04	0.93	-0.06
	Stabilized	349[a]	0.94	-0.04	0.93	-0.06
3	Unstabilized	349[a]	0.93	-0.04	0.93	-0.07
	Stabilized	349[a]	0.94	-0.04	0.93	-0.07
4	Unstabilized	359[c]	1.01	0.01	0.94	-0.05
	Stabilized	359[c]	1.01	0.01	0.94	-0.05
BDQ	Unstabilized	288[d]	0.96	-0.02	1.01	0.01
	Stabilized	288[d]	0.97	-0.02	1.02	0.02

Abbreviations: Sensitivity analysis restricted to participants also receiving bedaquiline (BDQ)

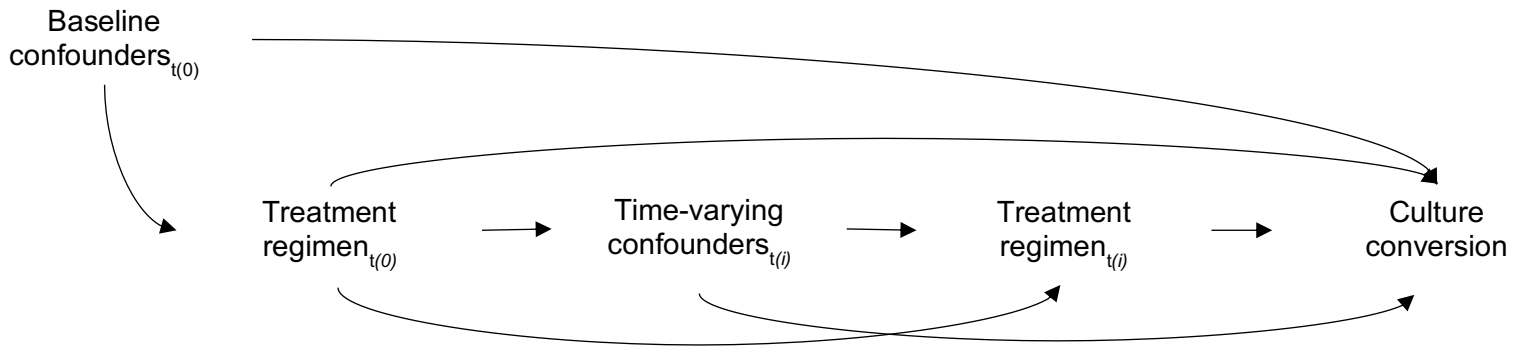
[a] N=10 participants excluded for missing time-varying data, N=2 participants excluded for missing baseline and time-varying data, N=2 participants excluded for missing baseline data

[b] Model used in primary analysis

[c] N=7 participants excluded for missing time-varying data N=1 participant excluded for missing baseline data

[d] N=4 participants excluded for missing baseline data

Appendix 5.4 Directed acyclic graph of baseline and time-varying confounding



Legend: The structure of bias assumed in Aim 1 is represented by the generalized directed acyclic graph above. We used content knowledge to select confounding variables. Baseline confounders used in the primary analysis include age, sex, whether the participant was in the hospital at treatment initiation, the number of Group A drugs in the regimen, whether the patient was on imipenem-cilastatin, body mass index <18.5, HIV infection, and hepatitis C; time-varying confounders used in the primary analysis include the number Group A drugs, sputum smear result, number of adverse events, and hospitalization.

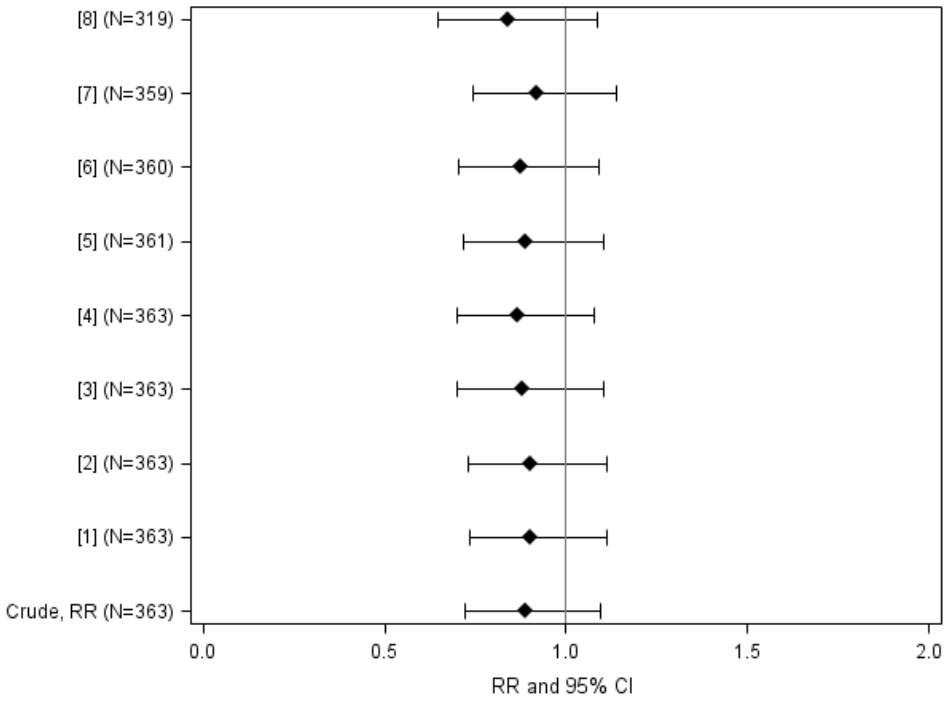
Appendix 5.5 Model specification for baseline confounders

We compared multiple models adjusted for baseline confounders. The below table indicates the variables included in each model on the corresponding forest plots for two-month and six-month culture conversion. Shading indicates the variable was included in the specified model.

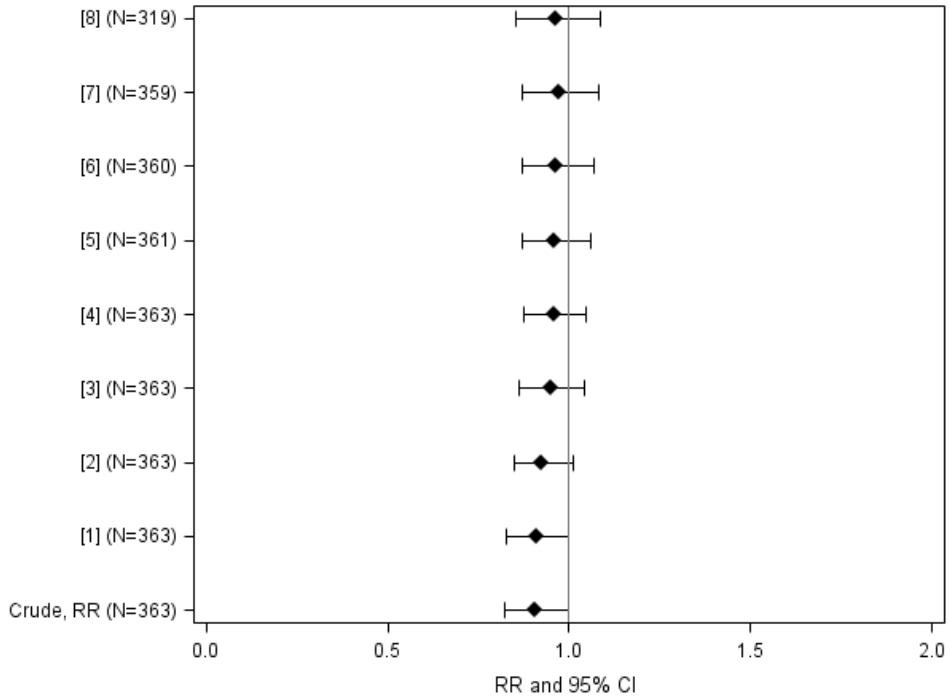
Covariate	Model number							
	1	2	3	4	5	6	7[a]	8
Age, sex	■							
In hospital at baseline	■	■						
# Group A drugs	■	■	■					
Imipenem/cilastatin	■	■	■	■				
Malnutrition	■	■	■	■	■			
HIV	■	■	■	■	■	■		
Hepatitis C	■	■	■	■	■	■	■	
Cavitation and smear grade ≥ 2++	■	■	■	■	■	■	■	■

[a] Model used in primary analysis

Two-month culture conversion



Six-month culture conversion



Appendix 5.6 Missing indicator analysis for baseline confounders

In Models 1-7 (Appendix 5.4), only 4/363 patients had missing data (N=2 for body mass index, N=1 for HIV, N=1 for Hepatitis C). Models using missing indicator variables for these covariates did not converge. Substantial data were missing for the composite variable for cavitary disease and smear grade (N=319/363) (Model 8, Appendix 5.4). We conducted a missing indicator analysis for the cavitary disease and smear grade variable following the principles of Greenland.(44) Results were nearly identical across analyses excluding the cavitation and smear variable from the model and including a missing indicator.

Analysis	Two-month culture conversion		Six-month culture conversion	
	N	RR (95% CI)	N	RR (95% CI)
No adjustment for cavitation and smear	359	0.92 (0.75, 1.14)	359	0.97 (0.87, 1.08)
Exclusion of patients missing data for cavitation and smear	319	0.84 (0.65, 1.09)	319	0.96 (0.85, 1.09)
Missing indicator for cavitation and smear	359	0.92 (0.74, 1.15)	359	0.97 (0.87, 1.08)

6 REFERENCES

1. Global tuberculosis report 2020. Geneva: World Health Organization; 2020. Licence: CC BY-NC-SA 3.0 IGO. Available at: https://apps.who.int/iris/rest/bitstreams/1312164/retrieve

2. Companion handbook to the WHO guidelines for the programmatic management of drug-resistant tuberculosis. (WHO/HTM/TB/2014.11) . Geneva, World Health Organization. 2015. Available at: http://apps.who.int/iris/bitstream/10665/130918/1/9789241548.

3. Bloss E, Kuk, a L, Holtz TH, Riekstina V, Skrip, et al. Adverse events related to multidrug-resistant tuberculosis treatment, Latvia, 2000-2004. International Journal of Tuberculosis and Lung Disease. 2010;14(3):275–281.

4. Cox HS, Kalon S, Allamuratova S, Sizaire V, Tigay ZN, Rüsch-Gerdes S, et al. Multidrug-Resistant Tuberculosis Treatment Outcomes in Karakalpakstan, Uzbekistan: Treatment Complexity and XDR-TB among Treatment Failures. PLoS One. 2007;2(11): e1126.

5. WHO treatment guidelines for drug-resistant tuberculosis, 2016 update (WHO/HTM/TB/2016.04). Geneva, World Health Organization. 2016. Available at: http://www.who.int/tb/areas-of-work/drug-resistanttb/treatment/resources/en/.

6. Global tuberculosis report 2018. Geneva: World Health Organization; 2018. License: CC BY-NC-SA 3.0 IGO. Available at: https://apps.who.int/iris/handle/10665/274453

7. The Use of Bedaquiline in the Treatment of Multidrug-Resistant Tuberculosis: Interim Policy Guidance. World Health Organization; 2013. Available at: https://apps.who.int/iris/rest/bitstreams/301963/retrieve

8. European Medicines Agency. European Medicines Agency recommends two new treatment options for tuberculosis. Press Release. 2013. https://www.ema.europa.eu/en/news/european-medicines-agency-recommends-two-new-treatment-options-tuberculosis

9. World Health Organization. WHO treatment guidelines for multidrug- and rifampicin-resistant tuberculosis, 2018 update. WHO/CDS/TB/2018.15. Available at: https://apps.who.int/iris/rest/bitstreams/1211676/retrieve

10. US National Library of Medicine. The Evaluation of a Standard Treatment Regimen of Anti-tuberculosis Drugs for Patients With MDR-TB (STREAM). ClinicalTrials.gov Identifier: NCT02409290. Available at: https://clinicaltrials.gov/ct2/show/NCT02409290?term=NC.

11. US National Library of Medicine. Evaluating the Safety, Tolerability, and Pharmacokinetics of Bedaquiline and Delamanid, Alone and in Combination, For Drug-Resistant Pulmonary Tuberculosis. ClinicalTrials.gov Identifier: NCT02583048. Available at: https://clinicaltrials.gov/ct2/show/NCT02583048

12. US National Library of Medicine. Evaluating Newly Approved Drugs for Multidrug-resistant TB (endTB). ClinicalTrials.gov Identifier: NCT02754765. Available at: https://clinicaltrials.gov/ct2/show/NCT02754765.

13. US National Library of Medicine. Safety and Efficacy of Various Doses and Treatment Durations of Linezolid Plus Bedaquiline and Pretomanid in Participants With Pulmonary TB, XDR-TB, Pre- XDR-TB or Non-responsive/Intolerant MDR-TB (ZeNix). Available at: https://www.clinicaltrials.gov/ct2/show/NCT03086486

14. World Health Organization. The use of delamanid in the treatment of multidrug-resistant tuberculosis interim policy guidance. Geneva: World Health Organization; 2014. Available at: https://apps.who.int/iris/handle/10665/137334

15. World Health Organization. Guidelines for the programmatic management of Multidrug-resistant Tuberculosis. Geneva: World Health Organization; 2011. Available at: http://whqlibdoc.who.int/publications/2006/9241546956_eng.pdf

16. World Health Organization. Rapid communication: key changes to treatment of multidrug- and rifampicin-resistant tuberculosis (MDR/RR-TB). Licence: CC BY-NC-SA 3.0 IGO; 2018. Available at: https://www.who.int/tb/publications/2018/WHO_RapidCommunicationMDRTB.pdf

17. Rubin DB. Estimating causal effects of treatments in randomized and nonrandomized studies. Journal of Educational Psychology. 1974;66(5):688–701.

18. Rubin DB. The design versus the analysis of observational studies for causal effects: parallels with the design of randomized trials. Statistics in Medicine. 2006/10/31. 2007;26(1):20–36.

19. Naimi AI, Cole SR, Kennedy EH. An introduction to g methods. International Journal of Epidemiology. 2017 Apr 1;46(2):756–762. Available from: https://www.ncbi.nlm.nih.gov/pubmed/28039382

20. Gler MT, Skripconoka V, Sanchez-Garavito E, Xiao H, Cabrera-Rivero JL, Vargas-Vasquez DE, et al. Delamanid for multidrug-resistant pulmonary tuberculosis. New England Journal of Medicine. 2012 Jun;366(23):2151–60.

21. von Groote-Bidlingmaier F, Patientia R, Sanchez E, Balanag VJ, Ticona E, Segura P, et al. Efficacy and safety of delamanid in combination with an optimised background regimen for treatment of multidrug-resistant tuberculosis: a multicentre, randomised, double-blind, placebo-controlled, parallel group phase 3 trial. The Lancet. Respiratory Medicine. 2019 Mar;7(3):249–259.

22. WHO position statement on the use of delamanid for MDR-TB. Licence: CC BY-NC-SA 3.0 IGO. Available at: https://www.who.int/tb/publications/2018/WHOPositionStatementDelamanidUse.pdf?ua=1.

23. Ahmad N, Ahuja SD, Akkerman OW, Alffenaar J-WC, Anderson LF, Baghaei P, et al. Treatment correlates of successful outcomes in pulmonary multidrug-resistant tuberculosis: an individual patient data meta-analysis. Lancet. 2018 Sep;392(10150):821–34.

24. Hafkin J, Hittel N, Martin A, Gupta R. Early outcomes in MDR-TB and XDR-TB patients treated with delamanid under compassionate use. European Respiratory Journal. 2017 Jul 1;50(1):1700311.

25. Ferlazzo G, Mohr E, Laxmeshwar C, Hewison C, Hughes J, Jonckheere S, et al. Early safety and efficacy of the combination of bedaquiline and delamanid for the treatment of patients with drug-resistant tuberculosis in Armenia, India, and South Africa: a retrospective cohort study. The Lancet. Infectious Diseases. 2018 May;18(5):536–44.

26. Auchynka V, Kumar AM V, Hurevich H, Sereda Y, Solodovnikova V, Katovich D, et al. Effectiveness and cardiovascular safety of delamanid-containing regimens in adults with multidrug-resistant or extensively drug-resistant tuberculosis: A nationwide cohort study from Belarus, 2016-18. Monaldi Archives for Chest Disease = Archivio Monaldi per le malattie del torace. 2021 Jan;91(1).

27. Ghosh S, Breitscheidel L, Lazarevic N, Martin A, Hafkin J, Hittel N. Compassionate Use of Delamanid in Adults and Children for Drug-resistant TB: 5-year Update. European Respiratory Journal. 2020 Jan 1;2002483.

28. Kwon Y-S, Jeon D, Kang H, Yim J-J, Shim TS. Concurrent Use of Bedaquiline and Delamanid for the Treatment of Fluoroquinolone-resistant Multidrug-resistant Tuberculosis: A Nationwide Cohort Study in South Korea. European Respiratory Journal. 2020 Jan 1;2003026.

29. Das M, Dalal A, Laxmeshwar C, Ravi S, Mamnoon F, Meneguim AC, et al. One step forward: Successful end-of-treatment outcomes of drug-resistant TB patients who received concomitant bedaquiline and delamanid in Mumbai, India. Clinical Infectious Diseases. 2020 Oct;ciaa1577. https://doi.org/10.1093/cid/ciaa1577

30. Khan U, Huerga H, Khan AJ, Mitnick CD, Hewison C, Varaine F, et al. The endTB observational study protocol: treatment of MDR-TB with bedaquiline or delamanid containing regimens. BMC Infectious Diseases. 2019 Aug;19(1):733.

31. World Health Organization. The use of bedaquiline in the treatment of multidrug-resistant tuberculosis interim policy guidance. Geneva: World Health Organization; 2013. Available at: https://apps.who.int/iris/rest/bitstreams/301963/retrieve

32. endTB Consortium. endTB Clinical and Programmatic Guide for Patient Management with New TB Drugs. Version 4.0; January 2018. Available at: https://samumsf.org/sites/default/files/2018-06/EndTB%20Guide%20for%20New%20TB%20Drugs%20Version%204.0.pdf

33. Lachenal N, Hewison C, Mitnick C, Lomtadze N, Coutisson S, Osso E, et al. Setting up pharmacovigilance based on available endTB Project data for bedaquiline. International Journal of Tuberculosis and Lung Disease. 2020 Oct;24(10):1087–1094.

34. García-Albéniz X, Hsu J, Hernán MA. The value of explicitly emulating a target trial when using real world evidence: an application to colorectal cancer screening. European Journal of Epidemiology. 2017 Jun;32(6):495–500.

35. Hernan MA, Robins JM. Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. American Journal of Epidemiology. 2016 Apr; 183(8):758–764.

36. Hernán MA, Sauer BC, Hernández-Díaz S, Platt R, Shrier I. Specifying a target trial prevents immortal time bias and other self-inflicted injuries in observational analyses. Journal of Clinical Epidemiology. 2016 Nov 1;79:70–75.

37. Labrecque JA, Swanson SA. Target trial emulation: teaching epidemiology and beyond. European Journal of Epidemiology. 2017;32(6):473–475.

38. Caniglia EC, Zash R, Jacobson DL, Diseko M, Mayondi G, Lockman S, et al. Emulating a target trial of antiretroviral therapy regimens started before conception and risk of adverse birth outcomes. AIDS. 2018 Jan;32(1):113–120.

39. Kurbatova E V, Cegielski JP, Lienhardt C, Akksilp R, Bayona J, Becerra MC, et al. Sputum culture conversion as a prognostic marker for end-of-treatment outcome in patients with multidrug-resistant tuberculosis: a secondary analysis of data from two observational cohort studies. The Lancet. Respiratory Medicine. 2015 Mar;3(3):201–209.

40. Meyvisch P, Kambili C, Andries K, Lounis N, Theeuwes M, Dannemann B, et al. Evaluation of six months sputum culture conversion as a surrogate endpoint in a multidrug resistant-tuberculosis trial. PLoS One. 2018 Jul 19;13(7):e0200539–e0200539.

41. Daniel RM, Cousens SN, De Stavola BL, Kenward MG, Sterne JAC. Methods for dealing with time-dependent confounding. Statistics in Medicine. 2013 Apr;32(9):1584–1618.

42. Rodriguez CA, Sy KTL, Mitnick CD, Franke MF. Time-Dependent Confounding in Tuberculosis Treatment Outcome Analyses: A Review of a Source of Bias. American Journal of Respiratory and Critical Care Medicine. 2020;202(9).

43. Hernán MA, Lanoy E, Costagliola D, Robins JM. Comparison of dynamic treatment regimes via inverse probability weighting. Basic & Clinical Pharmacology & Toxicology ; 2006; 98:237–242.

44. Greenland S, Finkle WD. A Critical Look at Methods for Handling Missing Covariates in Epidemiologic Regression Analyses. American Journal of Epidemiology. 1995 Dec 15;142(12):1255–1264.

45. Imperial MZ, Nahid P, Phillips PPJ, Davies GR, Fielding K, Hanna D, et al. A patient-level pooled analysis of treatment-shortening regimens for drug-susceptible pulmonary tuberculosis. Nature Medicine. 2018;24(11):1708–1715. Available from: https://doi.org/10.1038/s41591-018-0224-2

46. Franke MF, Khan P, Hewison C, Khan U, Huerga H, Seung KJ, et al. Culture Conversion in Patients Treated with Bedaquiline and/or Delamanid: A Prospective Multi-country Study. American Journal of Respiratory and Critical Care Medicine. 2020 Jul 24;

47. Diacon AH, Pym A, Grobusch MP, de los Rios JM, Gotuzzo E, Vasilyeva I, et al. Multidrug-Resistant Tuberculosis and Culture Conversion with Bedaquiline. New England Journal of Medicine. 2014 Aug 21;371(8):723–732.

48. Sanchez M, Bartholomay P, Arakaki-Sanchez D, Enarson D, Bissell K, Barreira D, et al. Outcomes of TB Treatment by HIV Status in National Recording Systems in Brazil, 2003–2008. PLoS One. 2012 Mar 21;7(3):e33129.

49. Young JG, Stensrud MJ, Tchetgen Tchetgen EJ, Hernán MA. A causal framework for classical statistical estimands in failure-time settings with competing events. Statistics in Medicine. 2020 Apr 15;39(8):1199–1236.

50. Wallis RS, Wang C, Meyer D, Thomas N. Month 2 culture status and treatment duration as predictors of tuberculosis relapse risk in a meta-regression model. PLoS One. 2013;8(8):e71116.

51. Wallis RS, Peppard T, Hermann D. Month 2 culture status and treatment duration as predictors of recurrence in pulmonary tuberculosis: model validation and update. PLoS One. 2015;10(4):e0125403.

52. Wallis RS, Peppard T. Early Biomarkers and Regulatory Innovation in Multidrug-Resistant Tuberculosis. Clinical Infectious Diseases. 2015;61Suppl 3:S160-3.

53. Phillips PPJ, Fielding K, Nunn AJ. An evaluation of culture results during treatment for tuberculosis as surrogate endpoints for treatment failure and relapse. PLoS One. 2013;8(5):e63840.

54. Davies GR. Early clinical development of anti-tuberculosis drugs: science, statistics and sterilizing activity. Tuberculosis. 2010;90(3):171–176.

55. Phillips PPJ, Mendel CM, Burger DA, Crook AM, Nunn AJ, Dawson R, et al. Limited role of culture conversion for decision-making in individual patient care and for advancing novel regimens to confirmatory clinical trials. BMC Medicine. 2016;14:19.

56. Bastard M, Sanchez-Padilla E, Hayrapetyan A, Kimenye K, Khurkhumal S, Dlamini T, et al. What is the best culture conversion prognostic marker for patients treated for multidrug-resistant tuberculosis? International Journal of Tuberculosis and Lung Disease. 2019;23(10):1060–1067.

57. Lu P, Liu Q, Martinez L, Yang H, Lu W, Ding X, et al. Time to sputum culture conversion and treatment outcome of patients with multidrug-resistant tuberculosis: a prospective cohort study from urban China. The European Respiratory Journal. 2017; 49(3):1601558.

58. Ahuja SD, Ashkin D, Avendano M, Banerjee R, Bauer M, Bayona JN, et al. Multidrug resistant pulmonary tuberculosis treatment regimens and patient outcomes: an individual patient data meta-analysis of 9,153 patients. PLoS Medicine. 2012;9(8): e1001300.

59. Siddique AA, Schnitzer ME, Bahamyirou A, Wang G, Holtz TH, Migliori GB, et al. Causal inference with multiple concurrent medications: A comparison of methods and an application in multidrug-resistant tuberculosis. Statistical Methods in Medical Research. 2018 Oct 31;28(12):3534–3549.

60. Mitnick CD, White RA, Lu C, Rodriguez CA, Bayona J, Becerra MC, et al. Multidrug-resistant tuberculosis treatment failure detection depends on monitoring

interval and microbiological method. European Respiratory Journal. 2016;48(4).1160–1170.

61. WHO consolidated guidelines on drug-resistant tuberculosis treatment. Geneva: World Health Organization; 2019. Licence: CC BY-NC-SA 3.0 IGO. Available at: https://apps.who.int/iris/rest/bitstreams/1211676/retrieve

62. Smith LH. Selection Mechanisms and Their Consequences: Understanding and Addressing Selection Bias. Current Epidemiology Reports. 2020;7:179–189.

63. Greenland S. Quantifying Biases in Causal Models: Classical Confounding vs Collider-Stratification Bias. Epidemiology. 2003;14(3).

64. Cole SR, Platt RW, Schisterman EF, Chu H, Westreich D, Richardson D, et al. Illustrating bias due to conditioning on a collider. International Journal of Epidemiology. 2010 Apr 1;39(2):417–420.

65. Whitcomb BW, Schisterman EF, Perkins NJ, Platt RW. Quantification of collider-stratification bias and the birthweight paradox. Paediatric and Perinatal Epidemiology. 2009 Sep;23(5):394–402.

66. Heinke D, Rich-Edwards JW, Williams PL, Hernandez-Diaz S, Anderka M, Fisher SC, et al. Quantification of selection bias in studies of risk factors for birth defects among livebirths. Paediatric and Perinatal Epidemiology. 2020 Nov 1;34(6):655–664.

67. Del Junco DJ, Bulger EM, Fox EE, Holcomb JB, Brasel KJ, Hoyt DB, et al. Collider bias in trauma comparative effectiveness research: the stratification blues for systematic reviews. Injury. 2015 May;46(5):775–780.

68. Mayeda ER, Glymour MM. The Obesity Paradox in Survival after Cancer Diagnosis: Tools for Evaluation of Potential Bias. Cancer Epidemiology, Biomarkers and Prevention. 2017 Jan 1;26(1):17–20.

69. Howe CJ, Cole SR, Lau B, Napravnik S, Eron Jr JJ. Selection Bias Due to Loss to Follow Up in Cohort Studies. Epidemiology. 2016 Jan;27(1):91–97.

70. Rudolph JE, Fox MP, Naimi AI. Simulation as a Tool for Teaching and Learning Epidemiologic Methods. American Journal of Epidemiology. 2020;190(5): 900–907

71. Lash TL, Fox MP, Fink AK. Applying Quantitative Bias Analysis to Epidemiologic Data. Springer Verlag, New York: 2009.

72. Rodriguez CA, Brooks MB, Aibana O, Mitnick CD, Franke MF. Reporting of definitions and analytic practices in multidrug-resistant tuberculosis cohorts using sputum culture-based endpoints. International Journal of Tuberculosis and Lung Disease. 2021; 25(7): 596-598.

73. World Health Organisation. Definitions and reporting framework for tuberculosis – 2013 revision. World Health Organisation. 2014. Available at: https://apps.who.int/iris/handle/10665/79199

74. Ndjeka N, Schnippel K, Master I, Meintjes G, Maartens G, Romero R, et al. High treatment success rate for multidrug-resistant and extensively drug-resistant tuberculosis using a bedaquiline-containing treatment regimen. European Respiratory Journal. 2018 Dec 1;52(6):1801528.

75. Conradie F, Diacon AH, Ngubane N, Howell P, Everitt D, Crook AM, et al. Treatment of Highly Drug-Resistant Pulmonary Tuberculosis. New England Journal of Medicine. 2020 Mar 4;382(10):893–902.

76. Gandhi NR, Shah NS, Andrews JR, Vella V, Moll AP, Scott M, et al. HIV Coinfection in Multidrug- and Extensively Drug-Resistant Tuberculosis Results in High Early Mortality. American Journal of Respiratory and Critical Care Medicine. 2010 Jan 1;181(1):80–86.

77. Schnippel K, Firnhaber C, Ndjeka N, Conradie F, Page-Shipp L, Berhanu R, et al. Persistently high early mortality despite rapid diagnostics for drug-resistant tuberculosis cases in South Africa. International Journal of Tuberculosis and Lung Disease. 2017 Oct 1;21(10):1106–1111.

78. Nahid P, Gonzalez LC, Rudoy I, de Jong BC, Unger A, Kawamura LM, et al. Treatment Outcomes of Patients with HIV and Tuberculosis. American Journal of Respiratory and Critical Care Medicine. 2007;175(11):1199–206.

79. Isaakidis P, Casas EC, Das M, Tseretopoulou X, Ntzani EE, Ford N. Treatment outcomes for HIV and MDR-TB co-infected adults and children: systematic review and meta-analysis. International Journal of Tuberculosis and Lung Disease. 2015 Aug;19(8): 969-978.

80. Hernan MA, Hernandez-Diaz S, Robins JM. A structural approach to selection bias. Epidemiology. 2004 Sep;15(5):615–625.

81. Toh S, Hernandez-Diaz S, Logan R, Robins JM, Hernan MA. Estimating absolute risks in the presence of nonadherence: an application to a follow-up study with baseline randomization. Epidemiology. 2010 Jul;21(4):528–539.

82. Keyes KM, Jager J, Platt J, Rutherford C, Patrick ME, Kloska DD, et al. When does attrition lead to biased estimates of alcohol consumption? Bias analysis for loss to follow-up in 30 longitudinal cohorts. International Journal of Methods in Psychiatric Research. 2020 Dec 1;29(4):e1842.

83. Teare MD, Dimairo M, Shephard N, Hayman A, Whitehead A, Walters SJ. Sample size requirements to estimate key design parameters from external pilot randomised controlled trials: a simulation study. Trials. 2014 Jul 3;15:264.

84. Molecular assays intended as initial tests for the diagnosis of pulmonary and extrapulmonary TB and rifampicin resistance in adults and children: rapid communication. Policy update. Geneva: World Health Organization; 2020. Licence: CC BY-NC-SA 3.0 IGO. Available at: https://apps.who.int/iris/handle/10665/330395

85. Maldonado G, Greenland S. Estimating causal effects. International Journal of Epidemiology. 2002 Apr;31(2):422–429.

86. Mansournia MA, Etminan M, Danaei G, Kaufman JS, Collins G. Handling time varying confounding in observational research. BMJ: British Medical Journal. 2017 Oct;359:j4587.

87. Hernán MA, Hernández-Díaz S. Beyond the intention-to-treat in comparative effectiveness research. Clinical Trials. 2011/09/23. 2012 Feb;9(1):48–55.

88. Toh S, Hernán MA. Causal inference from longitudinal studies with baseline randomization. International Journal of Biostatistics. 2008 Oct 19;4(1):22.

89. Greenland S. Interpretation and choice of effect measures in epidemiologic analyses. American Journal of Epidemiology. 1987 May 1;125(5):761–768.

90. Muller CJ, MacLehose RF. Estimating predicted probabilities from logistic regression: different methods correspond to different target populations. International Journal of Epidemiology. 2014 Jun 1;43(3):962–970.

91. McNutt L-A, Wu C, Xue X, Hafner JP. Estimating the Relative Risk in Cohort Studies and Clinical Trials of Common Outcomes. American Journal of Epidemiology. 2003 May 15;157(10):940–943.

92. Weuve J, Tchetgen Tchetgen EJ, Glymour MM, Beck TL, Aggarwal NT, Wilson RS, et al. Accounting for bias due to selective attrition: the example of smoking and cognitive decline. Epidemiology. 2012 Jan;23(1):119–128.

93. Koh W-J, Kwon OJ, Gwak H, Chung JW, Cho S-N, Kim WS, et al. Daily 300 mg dose of linezolid for the treatment of intractable multidrug-resistant and extensively drug-resistant tuberculosis. Journal of Antimicrobial Chemotherapy. 2009;64(2):388–391.

94. Udwadia ZF, Sen T, Moharil G. Assessment of linezolid efficacy and safety in MDR- and XDR-TB: an Indian perspective. European Respiratory Journal. 2010;35(4): 936–938–940.

95. Lee M, Lee J, Carroll MW, Choi H, Min S, Song T, et al. Linezolid for Treatment of Chronic Extensively Drug-Resistant Tuberculosis. New England Journal of Medicine. 2012;367(16):1508–1518.

96. De Lorenzo S, Alffenaar JW, Sotgiu G, Centis R, D'Ambrosio L, Tiberi S, et al. Efficacy and safety of meropenem-clavulanate added to linezolid-containing regimens in the treatment of MDR-/XDR-TB. European Respiratory Journal. 2013;41(6):1386–1392.

97. Sotgiu G, Centis R, D'Ambrosio L, Alffenaar J-WC, Anger HA, Caminero JA, et al. Efficacy, safety and tolerability of linezolid containing regimens in treating MDR-TB and XDR-TB: systematic review and meta-analysis. European Respiratory Journal. 2012 Dec;40(6):1430–1442.

98. Hernán MA. How to estimate the effect of treatment duration on survival outcomes using observational data. BMJ: British Medical Journal. 2018;360:k182.

CURRICULUM VITAE